

Untersuchung zur Beschreibung des diagnostischen Entscheidungsverhaltens von Psychiatern

Christine Tabbert-Haugg

1976



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41553810_0019

Aus der Nervenlinik der Universität München

Direktor: Professor Dr. H. Hippus

Untersuchung zur Beschreibung des diagnostischen
Entscheidungsverhaltens von Psychiatern

Inaugural-Dissertation
zur Erlangung der Doktorwürde in der gesamten Medizin
an der Ludwig Maximilians-Universität zu München

vorgelegt von
Christine Tabbert-Haugg
aus
München

1975

Mit Genehmigung des Fachbereichs Medizin
der Universität München

1. Berichterstatter: Prof. Dr. Hippius

2. Berichterstatter: Prof. Dr. Ploog

Dekan: Prof. Dr. Spann

Tag der mündlichen Prüfung: 24.2.1976



Herrn Professor Hippius danke ich für die Überlassung des Themas, seine Unterstützung und sein Verständnis, die das Entstehen dieser Arbeit ermöglicht haben.

Herrn Oberarzt Dr. von Cranach danke ich sehr herzlich für die Geduld und die hilfreiche Betreuung während der Durchführung dieser Arbeit.

Allen Ärzten, die aktiv an den diagnostischen Übungen teilgenommen haben, danke ich für ihre Bemühungen und ihre Ausdauer.

Inhaltsverzeichnis:

	Seite
Einleitung	1
Literatur	2
a) Longitudinalstudien	3
b) Studien zur Erfassung der Übereinstimmung von Psychiatern	4
c) Studien zur Erfassung der Übereinstimmung von Psychiatern verschiedener Länder (die internationale Übereinstimmung) in der Diagnostik psychiatr. Krankheitsbilder	10
d) Studien zum diagnostischen Prozess	13
Methodik	24
Zusammenfassung der Fragestellung	29
Ergebnisse	30
1. Diagnoseneinteilung	30
a) Beziehung zw. Anfangs- u. Enddiagnose	31
b) klinisch-psychiatr. Erfahrung und Wahl der Diagnose	40
2. Die subjektive Sicherheit in Abhängigkeit von der Diagnose	40
3. Die Persönlichkeit des Psychiaters	56
4. Die Symptombeobachtung in Abhängigkeit von der Wahl der Diagnose	59
5. Die Symptombeobachtung in Abhängigkeit von der klinisch-psychiatrischen Erfahrung der Beurteiler	61
6. Beschreibung des diagnostischen Entscheidungsverhaltens	64
a) Einteilung und Interpretation der sog. "Einzeldiagramme"	64
b) die Beziehung zwischen der endgültigen Diagnose und dem diagnostischen Verhalten eines Beurteilers	72
c) die Beziehung zwischen den Persönlichkeitseigenschaften der Psychiater und deren diagnostischem Verhalten	73
d) die Beziehung zwischen der klinisch-psychiatrischen Erfahrung der Beurteiler und deren diagnostischem Verhalten	73
e) Vergleich des diagnostischen Verhaltens von Beurteilern beim ersten und beim zweiten diagnostischen Interview	74

Diskussion

75

Zusammenfassung

84

Literaturverzeichnis

87

Anhang

EINLEITUNG

Ein Hauptproblem und schließlich auch eine der wichtigsten Aufgaben der wissenschaftlichen Tätigkeiten in der Psychiatrie ist die Erarbeitung von Methoden zur zuverlässigen und validen Erfassung psychopathologischer Merkmale, die dann schließlich zu einer psychiatrischen Diagnose führen. Dabei ist es von großer Wichtigkeit, die Ursachen für den bisherigen Mangel an Zuverlässigkeit psychiatrischer Diagnostik und Beobachtung aufzudecken.

Über die logische Begründbarkeit von Diagnosen im allgemeinen wurde viel diskutiert und geschrieben, diagnostische Reliabilität und Validität sind oftmals abgewogen und in Frage gestellt und eine Reihe mathematischer Modelle sind in diesem Zusammenhang für den Computergebrauch entwickelt worden. Verhältnismäßig umfangreiche Literatur über "Psycho-Diagnostik" wurde hauptsächlich in den 50er Jahren und früher von in Kliniken tätigen Psychologen hervorgebracht, die sich hauptsächlich mit der Ausarbeitung von Methoden zur Prüfung von Reliabilität und Validität der Diagnostik befaßten.

Allerdings hatten diese Arbeiten weniger Bedeutung für die diagnostische Praxis des Psychiaters im Sinne der herkömmlichen medizinischen Methodik, dem Explorieren und der Beurteilung von Symptomatik, Krankheitsätiologie und Prognose einer Erkrankung. Zahlreiche psychiatrische Studien und Schriften beklagten den Zustand mangelnder Reliabilität in der psychiatrischen Diagnostik und versuchten dies experimentell darzustellen. Doch erst später befaßte sich eine Reihe von Publikationen, insbesondere aus dem angelsächsischen Raum, mit den Ursachen dieser Probleme und der näheren Deskription des diagnostischen Prozesses in der Psychiatrie, sowie auch dem Aufdecken der Zuverlässigkeit diagnostischer Stile.

LITERATUR

Die Unzulänglichkeit der Reliabilität psychiatrischer Diagnostik war bisher von allen großen Reliabilitätsstudien auf dem Gebiet der Psychiatrie festgestellt und zum Teil stark kritisiert worden.

Geht man nun von der Definition aus, daß die Reliabilität einer Diagnose das Ausmaß aufzeigt, in dem die gleiche Diagnose bzw. diagnostische Kategorie nach der Wiederholung des gleichen diagnostischen Prozesses bei ein und dem selben Patienten wieder gestellt wird (Retest-Reliabilität), so ergeben sich zwei mögliche Gesichtspunkte:

1. die sog. "Langzeit"-Reliabilität (BECK 1962), d.h. die Änderung der Diagnose im Krankheitsverlauf, ausgedrückt durch das Ausmaß der Regelmäßigkeit, mit der der gleiche diagnostische Begriff für einen bestimmten Patienten zu einem festgelegten Zeitpunkt in verschiedenen späteren Stadien seiner Erkrankung, auch bei Änderungen seines klinischen Zustandsbildes, erneut angewandt wird.
2. Der andere Gesichtspunkt drückt sich darin aus, wie oft gleiche diagnostische Begriffe für einen bestimmten Patienten zu einem festgelegten Zeitpunkt seiner Erkrankung durch verschiedene Diagnostiker angewandt werden, d.h. wie groß die Übereinstimmung der Diagnostiker untereinander hinsichtlich der Diagnose eines bestimmten Patienten ist ("inter-rater Reliabilität").

Ein Klassifikationssystem also, das nur eine sehr geringe Übereinstimmungshöhe sogar unter gut ausgebildeten und erfahrenen Diagnostikern aufweist, wäre dementsprechend als nur wenig verlässlich zu bezeichnen. Für den Bereich der psychiatrischen Diagnostik kann die Reliabilität anhand der Übereinstimmungshöhe von verschiedenen Psychiatern, die denselben Patienten untersuchen, in Prozentsätzen bequem gemessen werden (KREITMANN 1961).

Nach diesen beiden Gesichtspunkten wurde nun dieses einfache Prinzip in jenen frühen Reliabilitätsstudien angewandt, die sich hauptsächlich mit der Frage der Verlässlichkeit der Diagnosen befaßten.

a) Longitudinal-Studien

Die Frage nach den Ursachen für die häufigen diagnostischen Veränderungen bei den gleichen Patienten, die in gewissen zeitlichen Abständen wiederholt in ein psychiatrisches Krankenhaus aufgenommen werden mußten, stellte ein wichtiges Problem beim Überprüfen der Verlässlichkeit psychiatrischer Diagnosen dar.

Wohl die bedeutendsten und einschlägigsten Arbeiten, die sich mit dieser Fragestellung befaßten, sind die Untersuchungen von einer Patientengruppe, deren Daten einer britischen Kohorten-Studie entnommen worden waren (COOPER, 1967). Bei den 293 Patienten, die innerhalb von zwei Jahren jeweils viermal in eine Nervenklinik in Großbritannien aufgenommen werden mußten, stellte sich eine überraschend hohe Anzahl von Diagnoseänderungen heraus. Nur 37 % der Patienten behielten während ihrer Wiederaufnahmen die anfangs gestellten Diagnosen unter Verwendung der dreistelligen ICD-Kategorien (International Classification of Diseases, Diagnoseschlüssel und Glossar psychiatrischer Krankheiten der WHO). Diese gefundene Ziffer sank unter Anwendung der vierstelligen ICD-Kategorien sogar auf 20 % der Fälle ab. Um nun die Gründe für diese Änderungen der Diagnosen zu untersuchen, wurden die jeweiligen Fallnotizen überprüft und ein ausgewählter Psychiater stellte für jeden einzelnen Patienten eine sog. "Standard-Diagnose". Der Vereinfachung wegen wurden die dreistelligen ICD-Kategorien nunmehr in acht diagnostische Gruppen eingeteilt (Schizophrenie, affektive Störungen, Süchte, nicht senile organische Störungen, senile Organstörungen und Verschiedenes). Die Ziffer der gleichbleibenden Diagnosen erhöhte sich hiermit auf 54 % der Fälle. Unter ausschließlicher Anwendung der sog. "Standard-Diagnosen" stieg diese Ziffer sogar auf 81 % der untersuchten Fälle an.

Aus diesen Ergebnissen schloß der Autor, daß die meisten Änderungen in den Klinikdiagnosen bei Wiederaufnahmen nicht auf einem Wandel des klinischen Zustandsbildes der entsprechenden Patienten beruhten, sondern daß vielmehr die eigentlichen klinischen Beobachtungen mit Hilfe einer standardisierten Methodik aufgezeichnet werden mußten, um verlässlichere Informationen über eben dieses klinische Zustandsbild des Patienten vermitteln und schließlich die Reliabilität von Krankenhausstatistiken erhöhen zu können (COOPER, 1967). MASSERMAN und CARMICHAEL (1938) verfolgten den Krankheitsverlauf von 100 Patienten nach ihrer Entlassung aus einem Nervenkrankenhaus. Sie stellten

in 40 % der Fälle eine Änderung der Diagnose anlässlich einer Nachuntersuchung nach einem Jahr fest. Die Frage der Reliabilität der Nomenklatur wurde hierbei zum ersten Male deutlich, wenn auch aus dem Bericht der Autoren hervorgeht, daß diese diagnostischen Veränderungen bei einem großen Teil der Fälle mit einem Wandel des klinischen Zustandsbildes der Patienten einhergingen.

Auch zwei weitere erwähnenswerte Untersuchungen (BARBIGAN, GARDNER, MILES und ROMANO, 1965; KAEHLING und VOLPE, 1963) befaßten sich mit der Problemstellung der Diagnoseänderungen bei Patienten mit mehreren stationären Aufenthalten in psychiatrischen Kliniken. In beiden Arbeiten wurden jeweils sehr viele Fälle überprüft und die erhaltenen Ergebnisse zeigten auch hier erhebliche diagnostische Veränderungen bei Patienten mit mehreren Wiederaufnahmen innerhalb weniger Jahre und auch diese Autoren glaubten die Gründe für diese gefundenen Phänomene diagnostischer Änderungen in erster Linie in dem Fehlen exakter diagnostischer Methoden und in einem mangelnden Wissen um die Ursachen der unterschiedlichen diagnostischen Gepflogenheiten der einzelnen Beurteiler zu sehen.

b) Studien zur Erfassung der Übereinstimmung von Psychiatern

In den folgenden Arbeiten setzte sich nun eine Anzahl von Autoren mit der Fragestellung auseinander, wie Psychiater im Diagnostizieren bestimmter Krankheitsbilder von Patienten übereinstimmen, die sie alle gleichzeitig, also in einem gewissen Stadium der Erkrankung, diagnostiziert hatten.

ASH (1949) überprüfte die Übereinstimmungshöhe mehrerer Psychiater hinsichtlich der diagnostischen Klassifikation von 52 männlichen Patienten eines staatlichen psychiatrischen Krankenhauses der U. S. A. . Die Patienten waren meist an einer sog. Charakter- oder Persönlichkeitsstörung erkrankt oder konnten als gesunde Persönlichkeiten bezeichnet werden, bei denen lediglich eine allgemeine Persönlichkeitsbeurteilung vorgenommen werden sollte. Die Interviews fanden in Form einer gemeinsamen Sitzung statt, während der jeder Psychiater jeweils die Fragen an den Patienten richten konnte, die er gerade für wichtig hielt. Die Übereinstimmung unter den Diagnostikern belief sich bei den sog. spezifischen Diagnosen (darunter fielen z. B. die schizoide Persönlichkeitsstruktur, konstitutionelle Minderwertigkeit, Drogenabhängigkeit, Epilepsie etc.) auf 32 % der Fälle, bei den sog. großen Kategorien (Psychosen,

Neurosen, Charakterstörungen allgemein) auf 64 % der untersuchten Fälle. Die gemeinsamen Interviews mit dem unvermeidbaren Gedankenaustausch unter den Diagnostikern, die nicht immer klar umrissene Nomenklatur, wie z.B. "primitive Persönlichkeit", und die etwas vagen Krankheitskategorien, die angewandt wurden und sich gegenseitig nicht immer unbedingt ausschlossen (z.B. Neurose und psychosomatische Störungen), erschwerten allerdings ein wirkliches Abschätzen der tatsächlich vorhandenen diagnostischen Übereinstimmung der Psychiater untereinander.

Die Häufigkeit der Diagnosenverteilung im Hinblick auf die Fragestellung, ob einzelne Psychiater bestimmte diagnostische Kategorien ganz allgemein bevorzugt anwenden, untersuchte MEHLMANN (1952) anhand von 4036 Patienten eines staatlichen Krankenhauses. Er überarbeitete hierbei die jeweiligen bereits vorhandenen Krankengeschichten dieser Patienten und versuchte somit, retrospektiv auf rein statistischer Grundlage festzustellen, wie häufig die einzelnen Psychiater bei diesen Patienten ganz bestimmte Krankheitsbilder diagnostizierten. Auf die einzelnen Krankheitssymptome, die zu den jeweiligen Diagnosen geführt hatten, ging der Autor allerdings nicht näher ein. Aus der Untersuchung ergab sich eine statistisch signifikante Variabilität unter den Psychiatern, die jedoch keinen Vergleich zuließ hinsichtlich der Verschiedenheit der Urteile einzelner Psychiater, die den gleichen Patienten diagnostiziert hatten.

Eine ähnliche Untersuchung mit entsprechenden Ergebnissen führten auch andere Autoren (PASAMANIK, DINTZ und LEFTON, 1959) durch. Die größten Differenzen, die sich aus dieser Studie ergaben, zeigten sich in der Diagnostik der Schizophrenie. Die Autoren folgerten aus den Ergebnissen, daß verschiedene Psychiater unterschiedliche Kriterien für das Diagnostizieren psychiatrischer Krankheitsbilder, insbesondere schizophrener und manisch-depressiver Psychosen, anwenden.

Die Untersuchung von RAINES und ROHRER (1954) befaßte sich mit 886 Offiziersanwärtern, deren Persönlichkeitsstruktur von einer Gruppe erfahrener Psychiater beurteilt werden sollte. Auch aus dieser Studie ergaben sich statistisch signifikante Unterschiede dahingehend, wie häufig bestimmte Psychiater - unabhängig von der beobachteten Symptomatik - gewisse Persönlichkeitsstörungen diagnostizierten. Ferner zeigte sich bei 124 Offiziersanwärtern, die jeweils von zwei verschiedenen Psychiatern unabhängig voneinander, also zweimal, exploriert wurden, jedesmal eine verschiedene Beurteilung ihrer Persönlichkeit.

HUNT, WITTON und HUNT (1953) verglichen die Diagnosen von 794 Marinesoldaten, die von Psychiatern der Marine voruntersucht und anschließend vom Dienst suspendiert wurden, mit den Diagnosen, die nach der Einweisung dieser Patienten in eine psychiatrische Klinik gestellt wurden. Die Klinikärzte hatten hierbei Zugang zu den Angaben der Psychiater von der Marine. Es zeigte sich eine allgemeine Übereinstimmung zwischen den beiden Institutionen von nur 32 % der Fälle bezüglich der spezifischen Diagnosen, während die Übereinstimmung in den weiter gefaßten Krankheitskategorien (Psychosen, Psychoneurosen, Persönlichkeitsstörungen) bei 54 % der Fälle lag. Ein beträchtlich höheres Maß an Übereinstimmung wurde in der Beurteilung der Tauglichkeit und der Bereitschaft zum Dienst erreicht (94 % der Fälle). Als eine besondere Schwierigkeit dieser Untersuchung bezeichneten die Autoren selbst u.a. auch den allzu großen Zeitintervall zwischen den beiden Beurteilungen der zwei Institutionen, sodaß sich das akute Krankheitsbild der Patienten habe erheblich wandeln und somit einen unmeßbaren Einfluß auf die jeweilige Diagnostik nehmen können.

Die Studie von SCHMIDT und FONDA (1956) belief sich auf 426 Patienten des Norwich State Hospital. Diese Patienten waren in der ersten Woche ihres stationären Aufenthaltes in dieser Klinik zunächst von einem jungen und relativ unerfahrenen Psychiater untersucht und diagnostiziert worden. Eine zweite Diagnose stellte dann jeweils in der dritten Woche nach der Klinikaufnahme ein sehr erfahrener und entschieden älterer Psychiater, der auch Zugang zu den Unterlagen und der Diagnose der ersten Untersuchung hatte.

Die Datenanalyse ergab eine allgemeine Übereinstimmungsquote dieser beiden Diagnostikerguppen, unter Ausschluß der organischen Fälle, von nur insgesamt 42 % der untersuchten Fälle hinsichtlich der spezifischen diagnostischen Kategorien. Besondere Merkmale dieser Arbeit sind die Anwendung der 1952 herausgegebenen Nomenklatur der American Psychiatric Association (APA) und nicht zuletzt auch der Versuch, wirksame Verbesserungsvorschläge für die statistischen Probleme dieser Art von Reliabilitätsstudien einzuführen.

Die Kritik an der Nosologie bewog schließlich auch FOULDS (1955) dazu, eine Untersuchung durchzuführen, die sich mit der Reliabilität psychiatrischer und der Validität psychologischer Diagnostik befaßte. In seiner etwas optimistischeren Anschauung gibt FOULDS selbst zu verstehen, daß die Ergebnisse "nicht ganz so spärlich" ausgefallen seien, "als es die derzeitige Meinung zu glauben verführt". Er hatte versucht, das Ausmaß an Übereinstimmung, das

von Psychiatern erreicht wurde, quantitativ unter Anwendung eines 6-Punkte-Systems zu bewerten. Dieses System stellte eine Skala dar, in der auch eine sog. Teilübereinstimmung in Betracht gezogen werden konnte. Die Skala basierte auf der Vorstellung, daß bestimmte Paare von Diagnosen, die einander ähnlich waren, z.B. "paranoider Zustand" und "paranoide Schizophrenie", eine höhere Bewertung bezüglich ihrer diagnostischen Übereinstimmung erhalten sollten, als Diagnosenpaare, die sich überhaupt nicht ähnlich waren, z.B. "paranoide Schizophrenie" und "Hysterie". Die Daten dieser Studie ergaben eine statistisch signifikante Übereinstimmung unter den beiden teilnehmenden Psychiatern für die 18 in dieser Versuchsreihe beurteilten Patienten. Mit Hilfe einer ähnlichen Untersuchungstechnik überprüfte FOULDS schließlich auch die Übereinstimmung von klinischen Psychologen in ihrer Diagnostik. Es zeigte sich dabei, daß klinische Psychologen, die den gleichen Patienten testen, unabhängig von den angewandten Testverfahren nicht besser übereinzustimmen scheinen, als die Psychiater in ihren diagnostischen Entscheidungen.

Eine sehr umfangreiche Untersuchung stellte schließlich auch NORRIS (1959) an. Die Autorin überprüfte die Diagnosen von über 6000 Patienten, die von Psychiatern ambulant eingehend untersucht, diagnostiziert und anschließend in eine Nervenklinik oder in ein geriatrisches Krankenhaus eingewiesen wurden. Dort wurden sie schließlich einige Wochen später nochmals diagnostiziert. Den zweiten Beurteilern, also den Klinikärzten, standen auch in dieser Untersuchung wesentlich mehr Informationen zur Verfügung, speziell die Diagnosen und eben die Kenntnis der Überweisungsstelle mit den vielleicht eigenen diagnostischen Gewohnheiten und nicht zuletzt die Krankengeschichten mit dem jeweiligen Krankheitsverlauf. Die einweisenden Psychiater, die ersten Beurteiler also, hatten keine derartigen Unterlagen zur Verfügung. Auch die jeweilige Zeitspanne zwischen der Erstuntersuchung, die meist zu Beginn einer floriden Erkrankung durchgeführt worden war, war in der Regel so groß, daß auch hier ein beträchtlicher Wandel des klinischen Zustandsbildes der Patienten möglich war. Beim Vergleichen der beiden Diagnosen fand nun die Autorin eine allgemeine Übereinstimmung von 60 % heraus, wobei jedoch organische Psychosen besonders stark vertreten waren.

KREITMAN (1961) wies auf die Notwendigkeit hin, die Reliabilität auch kleinerer Bestandteile und einzelner Faktoren, die zu der eigentlichen Diagnose führen, zu überprüfen. Eine detailliertere Analyse der verschiedenen diagnostischen Stile sollte somit angestrebt werden. In seiner Arbeit nämlich hatte er ver-

sucht, die Ergebnisse in der vorhandenen Literatur miteinander zu vergleichen, wobei sich herausstellte, daß sie nur wenig übereinstimmten. Er hielt somit solche Vergleiche auch nicht für besonders zweckmäßig. Er machte den Vorschlag, die Ursachen für diese Unterschiedlichkeiten in den einzelnen Untersuchungsergebnissen in fünf Hauptgruppen zu unterteilen, nicht zuletzt, um die Fehlerquellen für weitere Reliabilitätsuntersuchungen zu reduzieren und eine einheitlichere methodische Grundlage vorzubereiten. Diese fünf von KREITMAN vorgeschlagenen sog. Hauptgruppen bezogen sich einmal auf Verschiedenheiten in Hinblick auf die teilnehmenden Psychiater selbst (Alter, Erfahrung in der Psychiatrie, Art der Ausbildung, Interessen etc.), dann auf Unterschiede in der Untersuchungstechnik, in der angewandten Nomenklatur, in der Auswahl der Patienten und schließlich in der Art der Datenanalyse. Ferner wies er in seiner Publikation darauf hin, daß sich Reliabilitätsuntersuchungen nicht ausschließlich mit dem prozentualen Abschätzen einer allgemeinen Übereinstimmungsquote befassen sollten, sondern vielmehr auch mit den Ursachen der Meinungsverschiedenheiten in der Diagnostik psychiatrischer Krankheiten. Auch die Bedeutung und den Zweck der Anwendung diagnostischer Schemata hob der Autor besonders hervor. "Ein Fortschritt kann nicht aus der Betrachtung unscharfer Prozentsätze von Übereinstimmung kommen. Es ist weniger wertvoll zu wissen, daß soundsoviele Psychiater in einzelnen Diagnosen bei einem bestimmten Teil von Fällen übereinstimmen, als zu verstehen, warum sie in den verbleibenden Fällen nicht übereinstimmen" (KREITMAN, 1961).

In einer weiteren Studie berichten nun KREITMAN, SAINSBURY, MORRISSEY, TOWERS und SCRIVENER (1961) über ihre Untersuchung, die sie bereits unter der Berücksichtigung der Frage nach den Gründen für die so häufigen Uneinstimmigkeiten unter den psychiatrischen Diagnostikern hinsichtlich ihrer endgültigen Diagnosen durchgeführt hatten. Die von ihnen angewandte Methodik war auch hier wieder die Technik des sog. "Meinungspaarens", wobei 90 in das Chichester Mental Health Service neu aufgenommene Patienten von insgesamt 18 erfahrenen Psychiatern, zusammengestellt zu 9 Diagnostikerpaaren, innerhalb von drei Tagen untersucht wurden. Man achtete auf einheitliche Versuchsbedingungen für alle Teilnehmer, um eben die oben erwähnten Fehlerquellen so weit wie möglich zu vermeiden. Die Zeitintervalle zwischen den einzelnen Untersuchungen waren sehr kurz. Die Aufzeichnungen der Interviews und die angewandte Nomenklatur sind mit einiger Sorgfalt vorher gemeinsam bestimmt worden. Die Liste der Items enthielt Abschnitte über die

Dauer der Erkrankung, eine Checkliste von 28 Hauptsymptomen, Einzelheiten über vorhergehende Geisteskrankheiten, der Familiengeschichte und schließlich der Therapie des Patienten. Von 11 möglichen vorgeschlagenen Diagnosen durfte von den Beurteilern jeweils nur eine gestellt werden. Die Ergebnisse der beiden stets unabhängig voneinander durchgeführten Explorationen wurden anschließend verglichen und analysiert.

Die Autoren konnten die interessante Feststellung machen, daß insgesamt mehr Übereinstimmung in den endgültigen Diagnosen erzielt worden war, als in den Bewertungen einzelner Symptome. Obwohl die Psychiaterpaare hinsichtlich organischer Erkrankungen und endogener Psychosen sehr stark übereinstimmten, so war die Übereinstimmung in der Bewertung und auch Feststellung der Symptomatik, auf der diese Diagnosen basierten, wie z. B. das Auftreten von Intelligenzstörungen, Halluzinationen, Wahnleben, Aggression, Depression, Angst, Spannung etc., wesentlich geringer. Eine Differenzierung zwischen den Bestandteilen der vielen speziellen Urteile, die während einer Exploration gefällt werden, und den endgültigen diagnostischen Entscheidungen, die am Schluß den bestmöglichen Überblick verschaffen, sollte demnach nie außer Acht gelassen werden. Es gilt die Aussage: "Die Diagnose ist hauptsächlich eine Zusammenfassung vieler verschiedener Beobachtungen und als solche mit Wahrscheinlichkeit verlässlicher als irgendwelche dieser Bestandteile" (SHEPHERD, BROOKE, COOPER und LIN, 1968).

Entsprechende Ergebnisse brachte noch eine weitere Untersuchung (WING, BIRLEY, COOPER, GRAHAM und ISAACS, 1967). Ein strukturiertes Interview war von den Autoren vorbereitet worden, das die beiden an der Untersuchung teilnehmenden Psychiater bei einer bestimmten Anzahl von Patienten stets in Einzelsitzungen durchführten. Die allgemeine diagnostische Übereinstimmung erwies sich als sehr hoch, während Teilergebnisse bezüglich der Symptomatik, wie Depression, Angst, Spannung usw., wiederum eine überraschend niedrige Reliabilität zeigten.

Ein weiterer Schritt auf dem Gebiet der Reliabilitätsforschungen war somit unternommen worden. Wieder wurden Ideen und Hoffnungen wach, aus der Misere der unzulänglichen Reliabilität psychiatrischer Diagnostik allmählich herauszuwachsen. Irgendwann einmal sollte es sich herausstellen, daß, wenn man spezielle Schritte unternimmt, um die offensichtlichsten Quellen der Verschiedenheiten unter den psychiatrischen Beobachtern zu vermeiden, z. B. indem man standardisierte Interview-Richtlinien und allgemein anerkannte Glossars diagnostischer Terminologie anwendet, zufriedenstellendere Quoten von Reliabili-

tät und Wiederholbarkeit einzelner Stufen des diagnostischen Prozesses erzielt werden können (KENDELL, EVERETT, COOPER, SARTORIUS und DAVID, 1968; COOPER, 1970).

c) Studien zur Erfassung der Übereinstimmung von Psychiatern verschiedener Länder (die internationale Übereinstimmung) in der Diagnostik psychiatrischer Krankheitsbilder

Die unter Punkt b) angeführten Arbeiten befaßten sich wie diejenigen von Punkt c) mit der Fragestellung der Übereinstimmung von Psychiatern im Diagnostizieren, die folgenden Studien konzentrieren sich nun jedoch in erster Linie auf das immer deutlicher werdende Problem des unterschiedlichen diagnostischen Verhaltens von Psychiatern verschiedener Länder.

STENGEL (1959) vertrat z. B. die Ansicht, daß sich das Fehlen international vergleichbarer diagnostischer Schemata auf die Reliabilität der psychiatrischen Diagnose sehr nachteilig auswirke. Man erkannte die fundamentale Wichtigkeit epidemiologischer Untersuchungen, deren Gelingen wiederum nicht zuletzt auch von verlässlicheren und allgemein wiederholbaren Methoden für Diagnostik und Symptombewertung abhängt.

In Vergleichsstudien zwischen den Vereinigten Staaten und Großbritannien wurden alarmierende Ergebnisse registriert. Es stellte sich heraus, daß sich die diagnostischen Kriterien der Psychiater dieser beiden Länder - trotz der gemeinsamen Sprache - ganz erheblich voneinander unterscheiden. Dies wurde beispielsweise in Arbeiten, die im Rahmen des "U. S. - U. K., Diagnostic Project" durchgeführt worden waren, sehr deutlich (COOPER, KENDELL, GURLAND, SARTORIUS und FARKAS, 1969; COOPER, 1970; GURLAND, FLEISS, COOPER, SHARPE, KENDELL und ROBERTS, 1970). Es zeigte sich, daß die Psychiater in den staatlichen Krankenhäusern von New York viel häufiger eine schizophrene Erkrankung diagnostizierten als ihre Kollegen aus London.

Frühere Arbeiten anderer Autoren stellten ebenso unterschiedliche diagnostische Gepflogenheiten unter englischen und amerikanischen Psychiatern fest. SANDIFER, HORDERN, TIMBURY und GREEN (1968) benutzten in ihren Untersuchungen 30 gefilmte psychiatrische Interviews. Sie zeigten diese Filme jeweils einer Hörerschaft bis zu 10 Psychiatern in North Carolina, London und Glasgow. Die britischen Psychiater diagnostizierten hierbei zu einem wesent-

lich höheren Prozentsatz eine manisch-depressive Erkrankung als die amerikanischen Psychiater, umgekehrt verhielt es sich mit dem Diagnostizieren einer depressiven Neurose.

In einer weiteren Untersuchung mit entsprechender Methodik ergaben sich ähnliche Differenzen zwischen englischen und amerikanischen Psychiatern (SANDIFER, HORDERN, TIMBURY und GREEN, 1969).

KATZ, COLE und LOWERY (1969) stellten in ihrer Untersuchung fest, daß annähernd ein Drittel der amerikanischen Zuhörerschaft bei einem bestimmten, während der Exploration gefilmten Patienten die Diagnose einer schizophrenen Erkrankung stellten, ein weiteres Drittel der amerikanischen Psychiater bezeichnete ihn als neurotisch und die übrigen Beurteiler aus diesem Land diagnostizierten irgendeine Art von Persönlichkeitsstörung. Als der gleiche Film nun einer britischen Höhrerschaft im Maudsley Hospital, London, gezeigt wurde, stellte niemand die Diagnose einer Schizophrenie, aber 59 % der britischen Psychiater diagnostizierten eine Persönlichkeitsstörung, weitere 22 % eine Psychoneurose, 16 % schlugen eine "andere" Diagnose vor und 3 % stellten die Diagnose einer manisch-depressiven Erkrankung.

KENDELL, EVERETT, COOPER, SARTORIUS und DAVID (1968) und SHARPE, GURLAND, FISHER und FLEISS (1969) führten wieder im Rahmen des "U. S. - U. K. Project" Untersuchungen durch. Aus ihren Arbeiten ging hervor, daß sich die Psychiater der Vereinigten Staaten in ihrem Beurteilungsverhalten anhand von Rating-Scalen und eben auch den endgültigen Diagnosen sehr stark von ihren Kollegen aus Großbritannien unterscheiden. Am auffallendsten war dabei, daß die Psychiater aus den U. S. A. wesentlich mehr Merkmale des psychopathologischen Befundes bei Patienten wahrnahmen bzw. für pathologisch hielten, als dies bei den Psychiatern aus Großbritannien der Fall war. Sie neigten außerdem weit mehr dazu, eine schizophrene Erkrankung zu diagnostizieren.

Auch in zwei weiteren Arbeiten kamen die jeweiligen diagnostischen Gepflogenheiten amerikanischer und englischer Psychiater klar zum Ausdruck (COOPER, KENDELL, GURLAND, SHARPE, COPELAND und SIMON, 1972; KENDELL, COOPER, GOURLEY, COPELAND, SHARPE und GURLAND, 1971). Die englischen Psychiater stellten bevorzugt die Diagnose einer endogenen Depression, einer Manie, einer Neurose und einer Persönlichkeitsstörung. Die amerikanischen Psychiater hingegen diagnostizierten bei den gleichen Fällen bevorzugt eine schizophrene Erkrankung.

Zwischen Ländern also, die die gleiche Sprache sprechen und einen regen kulturellen und wissenschaftlichen Austausch untereinander pflegen, können sol-

che tiefgreifenden diagnostischen Verschiedenheiten bestehen. Dies bewog einige britische Autoren dazu, das diagnostische Verhalten von Psychiatern aus verschiedenen Gebieten Großbritanniens zu untersuchen (COPELAND, COOPER, KENDELL und GOURLEY, 1971). Etwa 200 Psychiater aus verschiedenen Teilen Großbritanniens nahmen an dieser Untersuchung teil. Sie beurteilten unter Anwendung von LORR's IMPS (Inpatient Multidimensional Psychiatric Rating Scale) drei gefilmte psychiatrische Interviews. Ihre Beurteilungsergebnisse und die Diagnosen wurden anschließend analysiert und miteinander verglichen. Es stellte sich eine überraschend gute allgemeine Übereinstimmung unter den Psychiatern heraus, lediglich die Psychiater aus Glasgow nahmen eine gesonderte Position ein. Sie neigten nämlich sehr stark dazu, eine affektive Erkrankung bei Patienten zu diagnostizieren, die von allen übrigen Psychiatern für schizophren gehalten wurden. Damit wurden auch die Ergebnisse einer früheren Untersuchung (SANDIFER, HORDERN, TIMBURY und GREEN, 1968) bestätigt.

Weitere Unterschiedlichkeiten im diagnostischen Verhalten von Psychiatern verschiedener Länder fand schließlich RAWNSLEY (1967) in seiner Arbeit heraus. Seine Ergebnisse beruhen auf niedergeschriebenen Krankengeschichten von insgesamt 30 Patienten. An der Untersuchung nahmen Psychiater aus Amerika, England, Dänemark, Norwegen und Schweden teil. Große Unterschiedlichkeiten ergaben sich hierbei selbst unter den Beurteilern der skandinavischen Länder.

Die Untersuchungen von PICHOT, BAILLY und OVERALL (1966) und von ENGELSMANN, VINAR, PICHOT, HIPPIUS, GIBERTI, ROSSI und OVERALL (1970) trugen dazu bei, diagnostische Stereotypen verschiedener Länder zu überprüfen. Erfahrene Psychiater aus Amerika, Frankreich, Deutschland, Italien und der Tschechoslowakei nahmen an diesen Projekten teil. Sie beurteilten jeweils mit Hilfe der B. P. R. S. (Brief Psychiatric Rating Scale, OVERALL und GORHAM, 1962) hypothetisch zusammengestellte Fallbeschreibungen von Patienten, die an einer endogenen Psychose erkrankt waren. Die Ergebnisse zeigten erfreulicherweise, daß die allgemeinen Begriffe der Schizophrenie, der Manie und der Depression von den Beurteilern dieser fünf Länder ähnlich verstanden wurden. Die Untergruppierungen dieser Krankheitsbegriffe und auch der anderen paranoiden und halluzinatorischen Psychosen, sowie auch deren jeweilige charakteristische Symptomatik wurden jedoch teilweise recht unterschiedlich verstanden.

Mit der Erforschung und Einführung der modernen Psychopharmaka-Therapie wurden nun auch quantifizierende Methoden zur psychiatrischen Befunderhebung zum wichtigen Bestandteil der Forschung auf diesem Gebiet. Die Wirkung der Therapie kann auf diese Weise genauer gemessen, objektiviert und auch allgemein verglichen werden. Aus diesem Grund wurde im Rahmen einer transkulturellen Studie eine Untersuchung durchgeführt, in der Patienten, die an einer akuten endogenen Psychose erkrankt waren, von den teilnehmenden Diagnostikern mit Hilfe der Bewertungsskala von LORR (IMPS, 1962) beurteilt wurden (BEHREND, FLEGEL, HELMCHEN, HIPPIUS, HÖFFKEN, SCHACHT und SCHULTE, 1971). Die in Deutschland faktorenanalytisch gewonnenen Ergebnisse oder Syndrome und Psychotikergruppen zeigten keine erwähnenswerten Unterschiede gegenüber den Resultaten der Vereinigten Staaten. Es konnten auf diese Weise nun unabhängig von irgendwelchen diagnostischen Klassifizierungen mit Hilfe von LORR's IMPS psychopathologische Phänomene bei endogenen Psychosen in verschiedenen Kulturkreisen miteinander verglichen werden.

d) Studien zum diagnostischen Prozess

Alle die oben beschriebenen Arbeiten mit ihren Ergebnissen und Beobachtungen hatten schließlich immer wieder die gleiche Frage aufgeworfen, nämlich welche Kriterien nun wirklich für die psychiatrische Untersuchung und das Diagnostizieren ausschlaggebend sind.

Die folgenden Studien befassen sich schwerpunktmäßig damit, die Reliabilität der weitgehend angewandten diagnostischen Stile aufzuzeichnen und vor allem die vielfältigen Einzelbestandteile eines diagnostischen Prozesses in der Psychiatrie zu erforschen.

Erstmals wies ZUBIN (1967) in seinen Überlegungen darauf hin, daß man über den Ursprung der diagnostischen Unterschiedlichkeiten in der Psychiatrie sehr wenig wisse. Die Untersuchungsergebnisse von KREITMAN (1961), der sich als erster damit beschäftigt hatte, dieser Fragestellung experimentell nachzugehen, waren von den Psychiatern ganz allgemein "recht pessimistisch" aufgenommen worden (SHEPHERD, BROOKE, COOPER und LIN, 1968).

Die erste große und wohl auch eine der bedeutendsten Studien, die letzten Endes damit anfängt, die Aufmerksamkeit direkt auf die Diagnostiker während ihrer Tätigkeit zu lenken, ist die von GAURON und DICKINSON (1966).

Das Hauptanliegen dieser Autoren bestand darin, zu erforschen, wie Psychiater zu einer diagnostischen Entscheidung kommen und bis zu welchem Ausmaß diese Entscheidungen auf verschiedenen Informations-Items basieren. "Diese Studie stellt einen Versuch dar, den entscheidungsfällenden Prozess in der Psychiatrie im ersten Brennpunkt zu erforschen, nämlich wie Psychiater eine Entscheidung treffen, ungeachtet dessen, ob diese "richtig" oder "falsch" sei (zitiert nach GAURON und DICKINSON, 1966). Insgesamt 12 Psychiater nahmen an dieser Studie teil. Sie wurden gebeten, in einer für sie gewohnten Art und Weise bestimmte Informationen, die aus Krankengeschichten gewonnen waren, von den Versuchsleitern zu erfragen. Diese Informationen stammten nun aus drei verschiedenen abgeschlossenen Krankengeschichten und waren in 40 verschiedene Kategorien gespalten worden. Die Reihenfolge, in der sie diese Informationen haben wollten, und auch eventuelle Zusatzfragen wurden den jeweiligen Beurteilern freigestellt. Jedem dieser 12 Psychiater wurde also eine Liste mit Überschriften der 40 Kategorien (wie z.B. Alter, Geschlecht, Familienanamnese, Persönlichkeitstestergebnisse etc.) vorgelegt. Nach jeder erfragten Information entsprechend dieser Überschriften wurden sie dann gebeten, die ihnen "als wahrscheinlich vorkommenden" Diagnosen mit dem jeweils subjektiv empfundenen Wahrscheinlichkeitsgrad in Prozent anzugeben. Sie sollten schließlich so lange nach weiteren Informations-Items verlangen, bis sie mit ihrer eigentlichen Diagnose zufrieden wären. Anschließend wurden die Psychiater gebeten, die Gründe für ihre endgültigen diagnostischen Entscheidungen mit ihren eigenen Worten wiederzugeben und die fünf für sie wichtigsten Informationen eigens zu vermerken. Die letzteren wurden als Kategorien "erkannter Wichtigkeit" (perceived category importance) bezeichnet. Die Kategorien "tatsächlicher Wichtigkeit" (actual category importance), also jene Kategorien, die nun aufgrund der Versuchsergebnisse erwiesenermaßen ihre diagnostischen Entscheidungen beeinflußt hatten, wurden den erstgenannten Kategorien gegenübergestellt. Es ergab sich, daß sich die Psychiater selbst meist nicht dessen bewußt waren, welche der Kategorien nun wirklich ihren diagnostischen Prozess am meisten beeinflußt hatten. Die Kindheitsgeschichte der Patienten stand z.B. an 16. Stelle unter "erkannter Wichtigkeit", jedoch an 6. Stelle unter "tatsächlicher Wichtigkeit". Entsprechend verhielt sich dies mit den übrigen Informationen. Die Psychiater stimmten allerdings verläßlich darüber überein, was für sie beim Abschätzen eines Falles, nachträglich gesehen, hätte wichtig sein sollen. Die notwendigsten Informationsteile, die zum Stellen einer endgültigen

Diagnose ausgereicht hatten, waren in ihrer Reihenfolge: der Grund für die Aufnahme des Patienten in die Klinik, dessen Alter, die prämorbidie Persönlichkeit, geistiger Zustand: Gedankeninhalt, Geschlecht, geistiger Zustand: allgemeine Erscheinung, frühere Erkrankungen, psychologische Testergebnisse, geistiger Zustand: Affekt, geistiger Zustand: Gedankenfluß.

Ohne nun direkt auf einzelne Daten einzugehen, führen die Autoren in ihrer Arbeit an, daß unerfahrenere Psychiater weniger Informationen erfragt, übereilter Entscheidungen getroffen und sich mehr auf den momentanen geistigen Zustand der Patienten verlassen hatten als ihre erfahreneren Kollegen. Schließlich zeigte sich noch, daß nur etwas mehr als die Hälfte der möglichen Informationen von der Gesamtheit der Beurteiler erfragt worden war, bis diese eine endgültige Diagnose stellen konnten.

GAURON und DICKINSON versuchten schließlich, in Anlehnung an eine Arbeit von MENDEL (1964), die einzelnen diagnostischen Stile gewissermaßen zu klassifizieren, indem sie das diagnostische Verhalten der einzelnen Psychiater in Dimensionen wie einer "strukturierten" entgegen einer "unstrukturierten" diagnostischen Methodik und einer "induktiv logischen" entgegen einer "intuitiv alogischen" Denkweise interpretierten.

Durch solche Beobachtungen, die die einzelnen Vorgänge während des diagnostischen Prozesses im Detail anführen, zeichnete es sich immer deutlicher ab, daß die Reliabilität der ständig kleiner gewordenen untersuchten Bestandteile des diagnostischen Verhaltens von Psychiatern entsprechend abnimmt, wenn man nun keine standardisierten und speziellen Richtlinien anwendet. Dieses Phänomen scheint schließlich zu immer größeren Uneinstimmigkeiten unter Beurteilern in ihrer diagnostischen Situation zu führen. "Es können die sonderbarsten und überraschendsten Beobachtungen zweier verschiedener Psychiater vorkommen, aber immer noch können sie in einer Diagnose übereinstimmen" (SHEPHERD, BROOKE, COOPER und LIN, 1968).

Nun herauszufinden, welche speziellen Merkmale eines Patienten in der Psychiatrie zu unzuverlässigen Beurteilungen führen können, drängt sich schließlich als weiteres Problem auf.

Könnten nun alle die erwähnten Ursachen, die zu diagnostischen Uneinstimmigkeiten führen, vermieden werden, so besteht dennoch die Notwendigkeit, eine gemeinsame diagnostische Terminologie zu schaffen, die jene Einzelbestandteile zusammenfügen soll.

BECK, WARD, MENDELSON, MOCK und ERBAUGH (1962) legten in ihrer Untersuchung besonders großen Wert darauf, einige Gründe für diagnostische Ver-

schiedenheiten in Hinblick auf die angewandten Klassifikationssysteme zu erfassen. Vier erfahrene Psychiater nahmen an den diagnostischen Übungen teil. Bevor sie sich mit den formalen Gesichtspunkten der Untersuchung beschäftigten, erörterten sie in vorhergehenden Sitzungen die verschiedenen diagnostischen Kategorien. Sie versuchten, Übereinstimmung bezüglich der spezifischen Kriterien der Nosologie in Anlehnung an das "Diagnostical and Statistical Manual" der "AMERICAN PSYCHIATRIC ASSOCIATION" zu erreichen. Die Psychiater wurden nun willkürlich zu Paaren zusammengestellt. Somit konnte jeder der 153 Patienten jeweils von zwei verschiedenen Beurteilern, die unabhängig voneinander zu ihren Diagnosen kamen, gesehen werden. Ein Psychiater explorierte also einen Patienten stets alleine, bis dieser nach einer Ruhepause von wenigen Minuten von dem zweiten Psychiater untersucht wurde. Danach trafen sich die Psychiater jeweils zur Erörterung ihrer Diagnosen. Die entstandenen Uneinstimmigkeiten wurden bei dieser Gelegenheit diskutiert. Die Gründe für die Meinungsverschiedenheiten bei insgesamt 40 Fällen wurden von den Beurteilern selbst in drei Hauptkategorien eingeteilt:

1. das nicht immer gleiche Verhalten des Patienten, das in 5 % der Fälle für die Uneinstimmigkeiten der Beurteiler verantwortlich gemacht wurde (2 Fälle),
2. das inkonstante Verhalten der Beurteiler selbst, verantwortlich in 32,5 % der Fälle (13 Patienten),
3. die Unzulänglichkeit der Nosologie, verantwortlich in 62,5 % der Fälle (25 Patienten).

Der Hauptgrund der Uneinstimmigkeiten unter den Psychiatern, für den sie sich also selbst verantwortlich glaubten, der oben angeführte Punkt 2, lag im verschiedenen Bewerten von Symptomen (7 Fälle), im unterschiedlichen "Material", das man aus der Exploration erfragt hatte (2 Fälle) und den unterschiedlichen Interpretationen der gleichen Pathologie (2 Fälle). Die Hauptuneinstimmigkeiten wurden jedoch in der Unzulänglichkeit des in der Studie angewandten APA-Systems gesehen, nämlich dem Zwang zu meist undurchführbar feinen diagnostischen Unterscheidungen. Diese drückten sich in Untergruppierungen aus, wie z.B. einer psychosomatischen Gestörtheit als Differentialdiagnose zu einer Persönlichkeitsstörung, die sich ja beide nicht gegenseitig ausschließen ließen.

Der Übereinstimmungsgrad unter den Psychiatern lag in dieser Studie bezüglich aller jener Diagnosen, derer sich die einzelnen Beurteiler sicher waren, bei 81 % der untersuchten Fälle und war damit wesentlich höher als in anderen etwa

vergleichbaren Reliabilitätsuntersuchungen. Was nun die Übereinstimmung der Beurteiler bezüglich der sog. spezifischen Diagnosen betraf, so ergab sich eine Ziffer von 54 % der entsprechenden Fälle. Auch diese Ziffer war größer als bei anderen Untersuchungen. Doch trotz der methodischen Verbesserungen in dieser Arbeit blieb das offensichtliche und grundlegende Problem der diagnostischen Meinungsverschiedenheiten weiter bestehen.

Der Mangel an 1. einheitlicher diagnostischer Terminologie und 2. einheitlichem diagnostischen Verhalten von Psychiatern wurde immer deutlicher.

Um dem Problem der unterschiedlichen Klassifikationen psychiatrischer Krankheitsbilder zu begegnen, überprüften schließlich die Autoren BOCHNIK, HELMCHEN, HIPPIUS, KNÜPPEL, KULENKAMPFF, LAUTER, MEYER, MÜLLER, WIESER und WINKLER (1970) die Brauchbarkeit der psychiatrischen Diagnosen der WHO-Diagnosen-Klassifikation (ICD, 8. Revision, 1970) an sechs deutschen psychiatrischen Krankenhäusern und Kliniken. Die Ergebnisse waren sehr befriedigend. Rund 90 % aller Patienten konnten bei der Anwendung von Diagnosen-Kombinationen, falls dies erforderlich war, gut charakterisiert werden. Der psychiatrische Teil der internationalen Klassifikation (ICD, 8. Revision), der sog. WHO-Diagnosenschlüssel, wurde nochmals in einer späteren Arbeit an vier Berliner Nervenkliniken untersucht (HELMCHEN, HIPPIUS und STÜRZBECHER, 1972). Auch hier zeigte es sich wieder, daß eine allgemeine Anwendung eines gemeinsamen Klassifikationssystems die allgemeine Übereinstimmungshöhe in der psychiatrischen Diagnostik wesentlich verbessern wird.

SANDIFER, PETTUS und QUADE (1964) hatten für ihre Reliabilitätsstudie eine Verfeinerung des Versuchsaufbaues angestrebt. Sie beabsichtigten, diesen den tatsächlichen Klinikbedingungen möglichst anzugleichen, um glaubwürdigere Verallgemeinerungen gegenüber den routinemäßig gesammelten Klinikdaten anstellen zu können. Es wurde in ihrer Untersuchung die sog. "Konferenzmethode" angewandt: zehn Diagnostiker, die in einer Gruppe zusammensaßen, diagnostizierten jeder für sich insgesamt 91 Patienten. Die Patientengruppe bestand aus Erstaufnahmen in ein öffentliches Nervenkrankenhaus und war im Alter von 15 bis 59 Jahren. Die Daten der Studie wurden in fünf diagnostischen Sitzungen über einen Zeitraum von sechs Monaten gesammelt. Unter Anwendung des APA-Systems ergab sich eine allgemeine Übereinstimmungsquote unter den Psychiatern bezüglich der endgültigen Diagnose von insgesamt 57 %. Somit konnte auch dieser Studie eine leichte Verbesserung der diagnostischen Reliabilität entnommen werden. Über den diagnostischen Vorgang selbst konnte jedoch keine besondere Aussage gemacht werden.

Eine andere Gruppe von Autoren (SHEPHERD, BROOKE, COOPER und LIN, 1968) versuchte schließlich herauszufinden, welche Rolle nun einzelne Faktoren des diagnostischen Prozesses in der Psychiatrie für die mangelnde Übereinstimmung unter den Psychiatern spielen.

In der ausführlichen Einleitung diskutierten die Autoren vier Hauptbestandteile einer psychiatrischen Untersuchung:

1. Die Interviewtechnik des Psychiaters, d.h. die Richtlinien, nach denen er während einer Exploration vorgeht, die Art und Weise, wie er bestimmte, gezielte Fragen an den Patienten richtet, in welcher Reihenfolge er seinen Interviewfragen nachgeht und inwieweit er den Patienten selbst gewissermaßen lenkt, ohne dessen persönliche Entfaltungsmöglichkeiten in der Fragenbeantwortung einzuengen.
2. Aus der Beantwortung der Interviewfragen und den verschiedenen Reaktionsweisen von seiten des Patienten ergeben sich schließlich ausführliche Informationen für den Psychiater. Die Autoren bezeichnen diesen Bestandteil des diagnostischen Prozesses eines Psychiaters als die "Wahrnehmung der Sprache und des Verhaltens des Patienten".
3. Aufgrund dieser Wahrnehmung von Sprache und Verhalten eines Patienten kommt es nun zu gewissen Folgerungen und diagnostischen Entscheidungen von Seiten des Psychiaters.
4. Als den letzten und endgültigen Schritt während eines diagnostischen Prozesses schilderten die Autoren die Wahl eines bestimmten diagnostischen Begriffes, also einer endgültigen Diagnose, gegenüber dem Patienten.

Diese strengen Unterteilungen halten die Autoren selbst zwar für künstlich, aber dennoch für notwendig, um diese Art von Reliabilitätsstudien sinnvoll durchführen zu können. Sie begründen dies mit dem verhältnismäßigen Mangel an äußeren Kriterien, wie z.B. EKG, Ergebnisse aus einem chemischen Labor usw., durch die die Entscheidungen und Beobachtungen des psychiatrischen Diagnostikers objektivierbar gemacht werden könnten.

Die eigentliche Untersuchung fand im Rahmen eines internationalen Seminars der WORLD HEALTH ORGANISATION statt. Einige Wochen vor Beginn dieser Tagung wurden den Teilnehmern dieser Untersuchung sechs Krankengeschichten von Patienten, die an einer endogenen Psychose erkrankt waren, zugeleitet. Bei der Auswahl der Patienten war darauf geachtet worden, einige

typische Fälle für Schizophrenie, endogener Depression und ebenso einige Fälle mit zweifelhafter Diagnose, hier z.B. einer schizoaffektiven Psychose, mit einzuschließen. Die Krankengeschichten enthielten Informationen über die Familienanamnese, die Eigenanamnese, die derzeitige Erkrankung mit dem körperlichen und dem psychischen Befund des Patienten und weitere spezielle Untersuchungsergebnisse aus der Klinik.

Die diagnostische Übung umfaßte insgesamt zwei getrennte Studien, eine mit den niedergeschriebenen Krankengeschichten und eine mit eigens für diese Arbeit vorbereiteten Fernsehaufzeichnungen.

Im ersten Teil der Untersuchung wurden also die Krankengeschichten bearbeitet. Unter Anwendung der ICD-Klassifikationen (5. Revision) wurden jeweils eine Hauptdiagnose und nach Belieben auch eine Zusatzdiagnose gestellt. Der jeweilige Sicherheitsgrad der angeführten Diagnosen, die in Frage kommenden Differentialdiagnosen und die eventuell noch benötigten Zusatzinformationen wurden eigens vermerkt. Die einzelnen Positionen der Krankengeschichten sollten von den Beurteilern selbst als jeweils "unwichtig", "wichtig" und "sehr wichtig" eingestuft werden. Dann notierten sie für jeden Patienten die Prognose und irgendwelche eigenen Bemerkungen zu den untersuchten Fällen.

Der zweite Abschnitt der Krankengeschichten beinhaltete den Krankheitsverlauf der Patienten. Die Psychiater wiederholten den gleichen Beurteilungsvorgang wie beim ersten Teil der Krankengeschichtsstudie, wobei eine Änderung der anfänglich gestellten Diagnosen gestattet war.

Die Fernsehaufzeichnungen der zweiten Studie bestanden aus neun psychiatrischen Interviews. Die Teilnehmer wurden gebeten, nach jedem dieser Interviews einen Fragebogen über den manifesten Befund des Patienten auszufüllen.

Über die Resultate und den allgemeinen Ausgang dieser diagnostischen Übungen konnte schließlich zusammenfassend bemerkt werden, daß eine experimentelle Annäherung an die tatsächlich vorhandene diagnostische Situation eines Psychiaters, die hier beispielsweise mit Hilfe der Fernsehfilme annähernd geschaffen wurde, am Ende dazu beitragen kann, einige Hauptbestandteile des diagnostischen Prozesses bis zum Erteilen einer klinischen Diagnose zu erfassen. Die Uneinstimmigkeiten und Kommunikationsschwierigkeiten unter den Psychiatern während der Studie schienen drei Hauptursachen zu haben:

1. unterschiedliche Beobachtungen und Wahrnehmungen der einzelnen Kliniker,

2. die vielfältigen Folgerungen, die aus allen jenen Beobachtungen gezogen wurden und
3. die unterschiedlichen Auffassungen hinsichtlich der diagnostischen Begriffe, die im Laufe einer solchen Studie zur Sprache kommen und der unterschiedliche Gebrauch diagnostischer Schemata von Seiten der einzelnen Psychiater verschiedener Schulen (SHEPHERD, BROOKE, COOPER und LIN, 1968).

Die Frage, inwieweit diese geschilderten Verschiedenheiten - sei es aufgrund der unterschiedlichen Persönlichkeiten der Psychiater oder aufgrund der verschiedenen diagnostischen Gepflogenheiten - nun die endgültig gestellten Diagnosen beeinflussen, blieb nach dieser Untersuchung jedoch weiterhin offen.

Eine neuere und sehr interessante Arbeit (SANDIFER, HORDERN und GREEN, 1970) befaßte sich hauptsächlich mit der Bestimmung des Zeitpunktes, zu dem diagnostische Entscheidungen gefällt bzw. die wichtigsten Symptome verzeichnet werden, die einen ausschlaggebenden Einfluß auf die endgültige Diagnose ausüben. Die Interviews von insgesamt 60 Patienten dauerten im Durchschnitt jeweils 25 Minuten. Sie wurden auf Fernsehfilme aufgenommen und Psychiatern im Alter zwischen 30 und 45 Jahren mit einer klinischen Erfahrung von durchschnittlich 10 Jahren vorgespielt. Die Psychiater stammten aus North Carolina, London und Glasgow. Sämtliche Filme wurden in genau festgelegten Zeitabständen während jeder Sitzung unterbrochen. Anschließend konnten die Psychiater bestimmte Bemerkungen hinsichtlich der erhaltenen Informationen auf ein Tonband sprechen. Schließlich wurden sie noch gebeten, eine Liste mit als "vorhanden" oder "nicht vorhanden" bezeichneten Symptomen auszufüllen und eine wahrscheinliche Diagnose zu stellen. In der darauffolgenden Phase der Untersuchung wurden die Filme nicht mehr unterbrochen. Jedoch gaben die Beurteiler jeweils die Anzahl der verstrichenen Minuten an, nach der sie bestimmte Beobachtungen gemacht hatten. Anschließend füllten sie eine 5-Punkte-Skala mit verschiedenen möglichen Diagnosen aus.

Die Analyse der Ergebnisse zeigte ein überraschendes Bild: die ersten drei Minuten einer diagnostischen Beobachtung hatten eine statistisch signifikante und damit wohl entscheidende Auswirkung auf die endgültige Diagnose (SHEPHERD, BROOKE, COOPER und LIN, 1968).

Eine weitere Variable des diagnostischen Prozesses in der Psychiatrie, nämlich die Auswirkung des persönlichen Kontaktes des Psychiaters zum Patienten - im Sinne einer direkten Gegenüberstellung von Arzt und Patienten - auf

die endgültige Diagnose, war bis dahin noch nicht näher untersucht worden. GAURON und DICKINSON (1969) setzten sich als erste Autoren mit dieser Problemstellung in einer Studie auseinander.

Die Fernsehaufzeichnungen der Studie beinhalteten zwei Patienteninterviews, die insgesamt 18 Psychiatern aus Iowa und Chicago vorgespielt wurden. Ein Teil der Psychiater hatte bereits vorher die Anamnese der Patienten zugesandt und damit die Möglichkeit bekommen, eine diagnostische Vorentscheidung zu treffen. Anschließend an die Filmvorführungen füllten sämtliche Psychiater einen Fragebogen (Patient Evaluation Form, SPIVAK, 1961) aus, wobei jedes der 26 angeführten Merkmale auf eine 5-Punkte-Skala eingetragen wurde. Dabei sollte jeder Beurteiler die subjektiv empfundene Wichtigkeit der einzelnen Merkmale eigens vermerken und einen Therapieplan vorschlagen.

Abschließend konnten die Autoren die Aussage machen, daß der psychopathologische Befund der Patienten, den die Psychiater aufgrund der eigenen Beobachtungen aus den Fernsehaufzeichnungen erhoben hatten, wesentlich mehr zu der endgültigen Diagnose beitrug, als irgendwelche Angaben aus den Anamnesen. Ferner zeigte sich mit statistischer Signifikanz, daß für das Diagnostizieren einer Psychose andere Items ausschlaggebend waren, als beispielsweise für das Diagnostizieren einer Persönlichkeitsstörung, wofür eher die jeweilige Ätiologie der Erkrankung von besonderer Wichtigkeit war.

Die größten Unterschiede bezüglich der endgültigen Diagnose ergaben sich zwischen den Psychiatern, die den Fernsehfilm bereits vor Erhalt der Anamnesen gesehen hatten, und jenen, denen zuerst die schriftlichen Informationen ausgehändigt worden waren, bis ihnen der Film gezeigt wurde.

Von großem Interesse ist schließlich auch jene Feststellung (die aufgrund der Studienergebnisse gemacht werden konnte), daß die meisten zur endgültigen Diagnose führenden Eindrücke fast bei allen Beobachtern innerhalb der ersten 30 bis 60 Sekunden des diagnostischen Prozesses gewonnen wurden. Dies wiederum bestätigt die Ergebnisse von SANDIFER, HORDERN und GREEN.

Eine der jüngsten Untersuchungen der diagnostischen Reliabilität in der Psychiatrie ist die Studie von KENDELL (1973). Eine Serie von 28 Fernsehaufzeichnungen kurzer und stets gleich strukturierter diagnostischer Interviews waren vorbereitet worden. In Anlehnung an die von SANDIFER (1970) und GAURON und DICKINSON (1969) angestellten Beobachtungen, nämlich daß die wichtigsten diagnostischen Beobachtungen meist sehr früh gemacht werden, dauerten die einzelnen Interviews jeweils nur fünf Minuten. Alle 28 willkürlich ausgesuchten Patienten waren Neuaufnahmen in das Maudsley Hospital, London.

Die teilnehmenden Psychiater verfügten über eine klinisch psychiatrische Erfahrung von mindestens vier Jahren. Der Inhalt der Interviews wurde auf drei verschiedenen Wegen wiedergegeben:

1. als vollständige Fernsehaufzeichnung mit Bild und Ton,
2. als Tonbandaufzeichnung alleine ohne Bild oder
3. nur in Form einer Abschrift des gefilmten Interviews, also ohne Bild und Ton.

Nach den ersten zwei Minuten der jeweils unterschiedlichen Wiedergaben sollten nun die Beurteiler eine Diagnose und eine Differentialdiagnose unter Anwendung der ICD-Nomenklatur (8. Revision) stellen. Die subjektive Sicherheit wurde auf einer Skala vermerkt. Anschließend wurden die Interviews auf dem gleichen Wege wie begonnen vollständig wiedergegeben. Die Diagnosen und subjektiven Sicherheiten führten die Beurteiler erneut an. Zusätzlich füllten sie noch eine Checkliste mit 55 Symptomen aus und vermerkten eigens diejenigen Items, auf denen ihrer Meinung nach ihre endgültigen Diagnosen jeweils beruhten.

Nach den ersten zwei Minuten der Interviews stimmten die Diagnosen der Beurteiler insgesamt mit den Klinikdiagnosen in ca. 50 % der Fälle überein. Nach 5 Minuten stieg diese Ziffer auf 64 % an. Die Übereinstimmung der Beurteiler untereinander betrug über 75 % der Fälle, und zwar unter Zugrundelegung aller drei Bewertungsbedingungen. Die allgemeine Klinikdiagnose wurde nun als Kriterium für die Richtigkeit der Diagnose angewandt. Es zeigte sich hierbei, daß sowohl die sog. Richtigkeit als auch die Übereinstimmung der Beurteiler untereinander für alle drei Versuchsbedingungen - vollständige Fernsehaufzeichnungen, Ton alleine bzw. die Interviewabschrift - gleich groß waren, trotz der Tatsache, daß in der Interviewabschrift keine Informationen über die äußere Erscheinung und das Verhalten des Patienten enthalten waren.

Aufgrund der Ergebnisse konnte nun folgende Aussage gemacht werden: Diagnosen, die von annähernd gleich ausgebildeten und entsprechend erfahrenen Psychiatern gestellt wurden, waren verhältnismäßig verläßlich. Ferner gilt, daß bei den meisten Patienten innerhalb der ersten fünf Minuten eines Interviews bereits eine endgültige Diagnose zustande kam, wobei das Verhalten und die äußere Erscheinung des Patienten offenbar keine besondere Rolle zu spielen schienen.

In Anbetracht der Vielfalt aller dieser Studien ist es nun schwierig, die meisten dieser Arbeiten miteinander zu vergleichen. Die Hauptgründe hierfür

liegen wohl in den sehr unterschiedlichen äußeren Experimentierahmen und den verschiedenen Kriterien für das Bewerten diagnostischer Übereinstimmung. Im allgemeinen zeigte es sich, daß es offenbar einfacher ist, Übereinstimmung in der Diagnostik von Psychosen zu erreichen, als im Feststellen einer neurotischen Erkrankung. Die Höhe der Übereinstimmung hängt zum größten Teil davon ab, wieviele diagnostische Kategorien angewandt werden bzw. wie fein diese unterteilt sind.

METHODIK

Das Ziel dieser Arbeit war es, mit Hilfe von gefilmten psychiatrischen Interviews die Zuverlässigkeit psychiatrischer Diagnostik zu untersuchen und herauszufinden, welche diagnostischen Überlegungen während einer psychiatrischen Exploration ablaufen, in welchem Stadium eines diagnostischen Interviews eine vorläufige Diagnose gewöhnlich erreicht wird und inwieweit dieser anfängliche Eindruck mit der endgültigen Diagnose am Schluß der Exploration übereinstimmt. Es sollte auch der Zusammenhang zwischen dem Verhalten eines Patienten und seinen tatsächlichen Informationen, die er durch die Beantwortung der Interviewfragen gibt, überprüft werden. Schließlich bestand das Ziel, die individuellen Unterschiede unter den Psychiatern hinsichtlich ihres diagnostischen Vorgehens und Verhaltens im Zusammenhang mit ihrer eigenen Persönlichkeit, ihrer klinisch psychiatrischen Erfahrung und dem Vertrauen, das sie für ihre jeweilige Diagnostik aufbringen, näher zu durchleuchten.

Es waren für die Untersuchung zwei Fernsehfilme von Patientenexplorationen vorbereitet worden. Die Filme hatten eine Gesamtdauer von 48 Minuten bzw. 44 Minuten und waren in regelmäßigen Zeitabständen, alle 4 Minuten im ersten Fall und alle 4,5 Minuten im zweiten Fall, mit einem Unterbrechungssignal versehen worden. Auf diese Weise sollten alle Psychiater, die an der Untersuchung teilnahmen, stets gleichlange Filmabschnitte sehen und damit für ihre jeweiligen Beurteilungen genau dieselben Informationen erhalten. Diese freien klinischen Interviews, die von einem erfahrenen Psychiater der Klinik, der die Patienten nicht kannte, durchgeführt worden waren, beabsichtigten, die Hauptsymptome des derzeitigen psychischen Zustandsbildes der Patienten und einen Teil der Anamnese hervorzuheben. Die Explorationen sollten keine kompletten Fallvorstellungen sein und keinem der Beurteiler wurden Zusatzinformationen gegeben.

Die teilnehmenden Ärzte waren vor Beginn der diagnostischen Übungen über Sinn und Ziel der Untersuchung informiert worden. Sie waren alle Assistenten oder Oberärzte der Universitäts-Nervenklinik München oder hatten früher dort ihre Facharztausbildung abgeschlossen und arbeiteten nun frei praktizierend. Mit der internationalen Klassifikation der WHO (ICD, 8. Revision, Psychiatrischer Teil, 1970) und dem AMP-System (Arbeitsgemeinschaft für Methodik

und Dokumentation in der Psychiatrie, 1972) waren sämtliche Teilnehmer vertraut. Die Mindesterfahrung der Ärzte in der Psychiatrie belief sich auf ein halbes Jahr klinischer Ganztagsbeschäftigung in der Universitäts-Nervenlinik München.

In Film 1 wirkten insgesamt 31 Beurteiler mit, in Film 2 waren es 35 Beurteiler. Ihre klinisch-psychiatrische Durchschnittserfahrung betrug 28,181 Monate (Standardabweichung: 22,474, Variationskoeffizient: 79,700, Standarderror: 2,766).

Bei der Auswahl der Patienten war darauf geachtet worden, daß die diagnostische Einordnung von Seiten des behandelnden Arztes bereits zu mehreren möglichen Differentialdiagnosen geführt hatte.

Beide Patienten waren kurz vor der Aufzeichnung des Interviews stationär in die Universitäts-Nervenlinik München aufgenommen worden. Beide hatten sich freiwillig für die Exploration und deren Aufzeichnung bereiterklärt. Den Beurteilern waren die Patienten sowohl namentlich als auch persönlich oder von deren Krankheitsgeschichte her völlig unbekannt.

Die für den ersten Film ausgewählte Patientin war zur Zeit der Exploration 23 Jahre alt. Sie stammte aus einem Landwirtschaftsbetrieb, hatte nach der Volksschule eine Realschule besucht und diese mit der mittleren Reife abgeschlossen. Ihre gesundheitlichen Beschwerden, wie z. B. "Kribbeln im Gesicht", "Druckgefühle im Hals mit Luftnot", "Herzklopfen" usw., seien vor 2 Jahren, nach einem Urlaub in Italien, erstmals aufgetreten. Vor einem Jahr sei bei ihr dann eine Kropfoperation vorgenommen worden, woraufhin sich ihr allgemeiner Zustand immer mehr verschlechtert habe. Die Patientin klagte schließlich über pelzige Füße und Arme, Steifigkeitsgefühle, so "einer Art Zahnschmerzen, die bis in die Hand- und Fußknochen" zu spüren gewesen seien.

Seit 6 Monaten glaubte die Patientin nun das Gefühl einer körperlichen Veränderung zu haben: ihr Körper sei ihr fremd geworden, insbesondere der Rücken sei ein anderer. Die Schulterblätter hätten sich nach oben verzogen, die Muskeln seien steif und einseitig verkürzt und im ganzen Körper spüre sie "so etwas ähnliches wie Ameisenlaufen". Mitunter käme es zu einem "Sausen" im ganzen Leib und auch zu starken Augenschmerzen mit zeitweiligen schweren Sehstörungen. Auch das Sprechen würde ihr sehr viele Beschwerden machen, alles sei "verlangsamt". Ganz besonders schlimm seien alle diese körperlichen Mißempfindungen unmittelbar nach Tabletteneinnahmen, hier wiederum besonders bei Einnahme der "Pille". Auch Alkohol und Zigaretten könne sie aus diesem Grunde nicht vertragen. Eine körperliche Erkrankung war nach Angaben der

Patientin ausgeschlossen worden. Sie selbst hielt eine "seelische Erkrankung" für möglich.

Auch bei der Klinikaufnahme hatten die von der Patientin geschilderten abnormen Leibgefühle im Vordergrund des psychopathologischen Bildes gestanden. Sie hatte auf die Aufnahmeärzte einen depressiv verstimmt und etwas antriebsarmen Eindruck gemacht. Ihr affektives Verhalten wurde als "läppisch" bezeichnet. Eine eindeutige Klassifizierung des Krankheitsbildes in Form einer endgültigen Klinikdiagnose war zum Zeitpunkt des Interviews noch nicht erfolgt. Die Diagnosen einer schizophrenen Erkrankung (insbesondere die coenästhetische Schizophrenie) und einer schweren neurotischen Erkrankung (hier insbesondere die hypochondrische Neurose) waren jeweils gleich stark in Erwägung gezogen worden.

Der für den zweiten Film ausgewählte Patient war 49 Jahre alt. Er berichtete in dem Interview, daß er von seiner Frau in die Klinik gebracht worden sei, nachdem er seit einigen Monaten in einem zunehmenden Glückszustand geraten sei, sich von den Lehren des "Guru Maharadshi" beflügelt fühle, nachts nicht mehr habe schlafen können und wahllos die Leute auf der Straße angesprochen habe, um über seinen Glückszustand zu berichten. In der Nacht vor der Einweisung sei er im Olympiagelände auf dem "Kalmarienberg" gewesen, habe dort sein ganzes Geld und seine Kleider weggeworfen, in der festen Überzeugung, jetzt in den Himmel zu fahren und dort "ja doch nichts mehr zu gebrauchen".

Der Patient berichtete weiter, er sei seit 2 Jahren an Meditationskursen beteiligt, habe hierdurch seine frühere "Gehemmtheit" überwunden und werde seit etwa sechs Wochen von Guru selbst beeinflusst. Man könne ihn jetzt nur noch zu irgendeiner wichtigen Handlung bewegen, wenn dies ein Befehl des Guru sei. Er, der Patient, sei schon immer sehr religiös und überaus gewissenhaft gewesen, was ihn schließlich zu einem guten Polizisten werden ließ. Er habe jedoch vor 7 Jahren so schwere Depressionen bekommen, daß er sich daheim in der Badewanne die Pulsadern aufschneiden wollte. Daraufhin sei er in das Nervenkrankenhaus in Haar eingewiesen worden und dort eineinhalb Jahre lang ohne jeden Lebenswillen geblieben. In den darauffolgenden Jahren, nach der Entlassung aus dem Nervenkrankenhaus Haar, sei er "mit den Nerven so schwach gewesen", daß er mehrmals den Beruf habe wechseln müssen. Seine größte Sorge sei nun, nachdem er sich wieder im Krankenhaus aufhalten müsse, wieder seine Stelle als Lagerarbeiter zu verlieren.

An dem vergangenen Pfingstsamstag, nach einem "Einweihungsritus", habe der

Patient schließlich ein "inneres Licht" gesehen, das verschiedenfarbig gewesen sei. Daraufhin habe er seine endgültige Glaubenssicherheit gewonnen. Stimmen habe er nie gehört. "Was ich tun und lassen soll, das geben mir mein Gewissen und mein Herz ein." Er glaubte auch, "etwas Besonderes" zu sein, indem Guru oft hilfeschuchende Menschen zu ihm geschickt habe, denen er "behilflich" sein sollte. Eigentlich krank sei er nicht, er leide lediglich unter Konzentrationsstörungen. Er sei immer täglich zur Arbeit gegangen, obwohl er nachts nur wenige Stunden habe schlafen können.

Der Patient bot für die behandelnden Klinikärzte ein sehr schwierig zu diagnostizierendes Krankheitsbild. Nicht zuletzt auch aufgrund seiner Vorgeschichte einigte man sich auf die Diagnose einer manisch-depressiven Psychose mit der Differentialdiagnose einer schizophrenen Erkrankung (paranoide Schizophrenie).

Dies war in etwa der Inhalt der beiden Explorationen. Die jeweiligen Symptome der Patienten sollten nun von den teilnehmenden Psychiatern beurteilt werden. Hierfür wurde das AMP-System angewandt. Die Beurteiler füllten nach jeder oben beschriebenen Filmunterbrechung jeweils das 3. Belegblatt des AMP-Systems (Psychischer Befund, 3. Psych. 3) und ein Formblatt mit der wahrscheinlichen Diagnose bzw. mehreren Differentialdiagnosen aus. Der erste Film hatte 11 Unterbrechungen, d.h. es wurden auch 11 voneinander völlig unabhängige Beurteilungen durchgeführt, beim zweiten Film waren es nur 9 Unterbrechungen und somit auch 9 Beurteilungen eben dieser verschiedenen Filmabschnitte. Die Psychiater waren gebeten worden, jeden einzelnen Filmabschnitt möglichst unabhängig von den bereits vorausgegangenen Informationsanteilen zu beurteilen. Sie durften auch die jeweils vorher von ihnen ausgefüllten Beurteilungsbögen und Formblätter nicht mehr einsehen.

Für jede der vorgeschlagenen Diagnosen bzw. Differentialdiagnosen waren jedesmal Angaben bezüglich des sog. subjektiven Sicherheitsgrades, d.h. der Wahrscheinlichkeit, die der Arzt der jeweils genannten Diagnose zumißt, in Prozent zu machen. Die Summe der Prozentangaben dieses subjektiven Sicherheitsgrades mußte immer gleich 100 betragen.

Das vorgeschriebene Klassifikationssystem für das Erteilen der Diagnosen war die WHO-Diagnosen-Klassifikation (ICD, 8. Revision, Psychiatrischer Teil, 1970). Die Diagnostiker waren gebeten worden, eine 4-stellige Klassifikation zu wählen.

Am Schluß der Untersuchung führten sämtliche Psychiater die sog. Halbform FPI-A des Freiburger Persönlichkeitsinventares (1970) durch. Auf einem an-

deren Formblatt machten sie schließlich noch Angaben über die Anzahl der Monate und Jahre ihrer klinischen Tätigkeiten in der Psychiatrie.

Um jeglichen Gedankenaustausch unter den Psychiatern bezüglich dieser Beurteilungen zu vermeiden, nahmen im Durchschnitt maximal zwei Ärzte gleichzeitig an einer diagnostischen Übung teil. Alle Teilnehmer wurden jedesmal gebeten, keinerlei Gespräche sowohl während als auch außerhalb einer Sitzung hinsichtlich ihrer Beobachtungen während der Fernsehaufzeichnungen zu führen. Ein gemeinsames Diagnostizieren bzw. das Erteilen irgendwelcher Zusatzinformationen für Beurteiler, die an der Untersuchung noch nicht teilgenommen hatten, sollte auf diese Weise vermieden werden.

Die einzelnen Filmabschnitte waren in Anlehnung an die Untersuchungen von SANDIFER, HORDERN und GREEN (1970) und GAURON und DICKINSON (1969) sehr kurz bemessen.

Die Sitzungen waren für die teilnehmenden Psychiater sehr anstrengend und langwierig. Jeder Beurteilungsvorgang dauerte pro Film insgesamt zwischen zwei und vier Stunden.

ZUSAMMENFASSUNG DER FRAGESTELLUNG

Mittels der unter "Methodik" geschilderten Versuchsanordnung wollten wir gewisse Teilbereiche des diagnostischen Entscheidungsverhaltens von Psychiatern näher beschreiben. Dieses Psychiaterverhalten wird im nächsten Kapitel der Ergebnisse behandelt.

Dazu gehört als erstes der Vergleich zwischen den ursprünglichen diagnostischen Entscheidungen von Psychiatern mit deren Abschlußdiagnosen am Ende der Untersuchung, womit die Beziehung zwischen der Anfangs- und der Enddiagnose überprüft werden sollte.

Anschließend suchten wir nach Zusammenhängen zwischen der klinisch-psychiatrischen Erfahrung von Psychiatern und deren diagnostischen Entscheidungen.

In einem weiteren Schritt überprüften wir, ob die sog. subjektive Sicherheit der Beurteiler unserer Interviews irgendwelche Einflüsse auf deren Diagnosenwahl ausübte.

Die Persönlichkeit der Psychiater und deren diagnostisches Entscheidungsverhalten, die Symptombeobachtungen und die Wahl der Diagnosen, die klinisch-psychiatrische Erfahrung der einzelnen Beurteiler und deren Symptombeobachtungen wurden schließlich jeweils in Abhängigkeit voneinander untersucht.

Als nächstes wollten wir das diagnostische Entscheidungsverhalten von Psychiatern näher beschreiben, indem wir es graphisch darstellten. Dies versuchten wir mit Hilfe der sog. "Einzeldiagramme", die wir interpretierten und in zwei Hauptgruppen einteilten. Das auf diese Weise gefundene diagnostische Verhalten der einzelnen Psychiater setzten wir dann in Beziehung zu deren endgültig gewählten Diagnosen, zu deren Persönlichkeitseigenschaften, zu deren Symptombeobachtungen und schließlich zu deren klinisch-psychiatrischer Erfahrung.

Am Schluß verglichen wir das jeweilige diagnostische Verhalten der Beurteiler des ersten und des zweiten diagnostischen Interviews, mit der Fragestellung, ob die Psychiater dieses diagnostische Verhalten von Fall zu Fall beibehalten oder ändern.

ERGEBNISSE

1. Die Diagnoseneinteilung

Bei der Auswertung der Untersuchungsergebnisse wurden zunächst die drei- und vierstelligen ICD-Diagnosen wegen der besseren Übersicht in mehrere diagnostische Hauptgruppen, entsprechend ihrer klassifikatorischen Ähnlichkeit, eingeteilt.

Bei Film 1 ergaben sich entsprechend den von den Beurteilern angebotenen Diagnosen fünf verschiedene diagnostische Hauptgruppen:

Die 1. Hauptgruppe umfaßt alle Neurosen, Persönlichkeitsstörungen (Psychopathien) und alle anderen nicht psychotischen psychischen Störungen, die in dem WHO-Diagnosenschlüssel unter den drei- und vierstelligen ICD-Nummern von 300 (Neurosen) bis einschließlich Nr. 305 (Psychosomatische Störungen) angegeben sind.

Die 2. diagnostische Hauptgruppe umfaßt die affektiven Psychosen (ICD-Nr. 296) und die äußerst selten in Erwägung gezogene reaktive depressive Psychose (ICD-Nr. 298.0).

In der 3. Hauptgruppe wurden alle Formen von Schizophrenie, also alle Unterformen der dreistelligen ICD-Nr. 295 zusammengefaßt.

Die 4. Hauptgruppe umfaßt alle körperlich begründbaren psychiatrischen Krankheitsbilder. Hierunter fallen die "Psychosen bei anderen körperlichen Krankheiten" (ICD-Nr. 294), speziell "bei endokrinen Störungen" (ICD-Nr. 294.0) und "bei Intoxikationen durch Arzneimittel oder Gifte" (ICD-Nr. 294.3).

Ferner enthält diese Hauptgruppe die "Psychischen Störungen, die nicht als Psychosen bezeichnet werden können, jedoch mit körperlichen Krankheiten im Zusammenhang stehen, die auf das Gehirn einwirken" (ICD-Nr. 309). Schließlich wurden noch die "Neurologischen Krankheiten" (ICD-Nr. 316.1) in diese Hauptgruppierung mit hineingenommen.

Alle Klassifikationen dieser 4. Hauptgruppe wurden äußerst selten sowohl im ersten als auch im zweiten Film diagnostiziert, sodaß eine Trennung zwischen psychotischen und nicht psychotischen psychiatrischen Krankheitsbildern als Folge einer organischen Erkrankung der Übersicht halber nicht vorgenommen wurde.

Die 5. diagnostische Hauptgruppe beinhaltet nun noch die "psychiatrisch unklaren Fälle" (ICD-Nr. 316.0), die am seltensten als Diagnose in Erwägung gezogen wurden.

Bei Film 2 ergaben sich schließlich sechs diagnostische Hauptgruppen, die wiederum entsprechend den angebotenen Diagnosen und ihrer klassifikatorischen Ähnlichkeit eingeteilt wurden. Diese Hauptgruppen entsprechen im Wesentlichen denen des ersten Filmes mit folgenden Ausnahmen:

Zur 3. Hauptgruppe von Film 1 (alle Schizophrenien, ICD-Nr. 295) kam lediglich die sehr selten diagnostizierte "Paranoide Psychose im Involutionalter" (ICD-Nr. 297.1) hinzu.

Die 4. Hauptgruppe nach Film 1 umfaßt in Film 2 als weitere Krankheitsbilder die "Psychosen bei intrakraniellen Störungen" (ICD-Nr. 292) und die "Psychosen bei anderen organischen Hirnstörungen" (ICD-Nr. 293).

Eine 6. Hauptgruppe wurde nun zusätzlich noch aufgestellt für die "Schizoaffective Psychose" (ICD-Nr. 295.7), die in Film 2 sehr häufig als Differentialdiagnose gegenüber den Schizophrenien, die die übrigen Krankheitsbilder der ICD-Nr. 295 ausmachen, gestellt wurde.

a) Beziehung zwischen Anfangs- und Enddiagnose

Als erstes untersuchten wir, wieviele Diagnostiker ihre ursprüngliche diagnostische Entscheidung nach dem ersten Informationsabschnitt des gefilmten Interviews im Laufe ihres diagnostischen Prozesses bis zum Schluß der Untersuchung beibehielten bzw. änderten.

Die "wahrscheinlichsten Diagnosen" nach jedem Informationsanteil (IA), d. h. nach jedem Filmabschnitt, waren jeweils diejenigen Diagnosen, die von den Beurteilern mit dem höchsten Sicherheitsgrad in Prozent (SG%) angegeben worden waren. Als "endgültige Diagnose" wurde die Diagnose mit dem höchsten subjektiven Sicherheitsgrad am Ende des diagnostischen Interviews (bei Film 1 also nach dem 11. Informationsanteil, bei Film 2 nach dem 9. Informationsanteil) bezeichnet.

In den folgenden zwei Tabellen sind nun die Diagnosen mit dem jeweils höchsten

Sicherheitsgrad der einzelnen Beurteiler während der gesamten diagnostischen Übungen bis hin zur jeweils "endgültigen Diagnose" (bei Film 1 nach dem 11. IA, bei Film 2 nach dem 9. IA) in Abhängigkeit von der Gesamtzahl der Beurteiler in Prozent angegeben. Es gelten die oben beschriebenen diagnostischen Hauptgruppen.

	1. IA	2. IA	3. IA	4. IA	5. IA	6. IA	7. IA	8. IA	9. IA	10. IA	11. IA
1. H. Gr. Beurteiler in %	23 %	20 %	18 %	15 %	13 %	17 %	16 %	16 %	18 %	18 %	19 %
2. H. Gr. Beurteiler in %	7 %	5 %	5 %	8 %	11 %	8 %	8 %	8 %	8 %	8 %	7 %
3. H. Gr. Beurteiler in %	-	5 %	7 %	7 %	6 %	5 %	6 %	6 %	5 %	5 %	5 %
4. H. Gr. Beurteiler in %	-	-	-	-	-	-	-	-	-	-	-
5. H. Gr. Beurteiler in %	3 %	3 %	3 %	3 %	3 %	3 %	3 %	3 %	-	-	-

Abb. Nr. 1.: Film 1. Prozente der Beurteiler, deren Hauptdiagnosen (höchster subjektiver Sicherheitsgrad) in Abhängigkeit von den einzelnen Informationsabschnitten dargestellt sind.

IA = Informationsanteil

1. H. Gr. = ICD-Nr. 300 - 305

2. H. Gr. = ICD-Nr. 296, 298

3. H. Gr. = ICD-Nr. 295

4. H. Gr. = ICD-Nr. 292, 293, 294, 316.1

5. H. Gr. = ICD-Nr. 316.0

6. H. Gr. = ICD-Nr. 295.7

	1. IA	2. IA	3. IA	4. IA	5. IA	6. IA	7. IA	8. IA	9. IA
1. H. Gr. Beurteiler in %	37 %	23 %	11 %	11 %	16 %	14 %	14 %	8 %	6 %
2. H. Gr. Beurteiler in %	-	3 %	3 %	5 %	5 %	6 %	9 %	16 %	23 %
3. H. Gr. Beurteiler in %	46 %	63 %	72 %	70 %	65 %	66 %	17 %	18 %	19 %
4. H. Gr. Beurteiler in %	3 %	-	3 %	3 %	3 %	3 %	3 %	-	-
5. H. Gr. Beurteiler in %	6 %	-	-	-	-	-	-	-	3 %
6. H. Gr. Beurteiler in %	8 %	11 %	11 %	11 %	11 %	11 %	26 %	23 %	14 %

Abb. Nr. 2: Film 2. Prozente der Beurteiler, deren Hauptdiagnosen in Abhängigkeit von den einzelnen Informationsabschnitten dargestellt sind.

Die Prozentangaben der vorangehenden Abbildungen drücken aus, wieviele Beurteiler pro Informationsabschnitt (IA) eine bestimmte diagnostische Hauptgruppe als die "wahrscheinlichste Diagnose" bzw. die "endgültige Diagnose" jeweils angegeben haben. Anhand des verhältnismäßig gleichförmigen Verlaufs der Zahlenbilder vom 1. Informationsabschnitt der Fernsehaufzeichnungen bis hin zum letzten Informationsabschnitt (11. IA bei Film 1, 9. IA bei Film 2) läßt sich erkennen, daß in etwa stets gleichviele Beurteiler eine gegebene diagnostische Hauptgruppe als die "wahrscheinlichste Diagnose" oder die "endgültige Diagnose" bevorzugt diagnostiziert haben.

Die Ergebnisse von Film 2 weichen hierbei etwas von denen des ersten Filmes ab, ohne jedoch die Aussage zu widerlegen. Ein gewisser Trend der Gleichförmigkeit läßt sich auch anhand der Ergebnisse des zweiten Filmes feststellen.

Zur Übersicht seien nun noch in einer Tabelle alle diejenigen Diagnosen zusammengestellt, die von den Beurteilern jeweils am Ende des diagnostischen Interviews als die "endgültige Diagnose" bezeichnet wurden. Diese sog. "endgültige Diagnose" stellte also diejenige Diagnose dar, die von den einzelnen Ärzten jeweils mit dem größten subjektiven Sicherheitsgrad (SG%) am Ende der Untersuchung versehen worden war. Es sei darauf hingewiesen, daß auch am Ende der diagnostischen Übungen die Möglichkeit bestand, mehrere Differentialdiagnosen, jedoch keine Zusatzdiagnosen (!), mit einem entsprechend niedrigeren subjektiven Sicherheitsgrad anzugeben.

Bei dieser Übersicht wurde nun auf die Einteilung der einzelnen Diagnosen in diagnostische Hauptgruppen verzichtet. Alle diagnostischen Entscheidungen wurden als vierstellige ICD-Nummern bzw. deren Bezeichnung wiedergegeben. Die Prozentangaben beziehen sich auf die jeweilige Anzahl der Psychiater, die diese Diagnose als "endgültige Diagnose" gewählt hatten. Diese beiden Angaben, Prozentsatz der Diagnostiker und deren "endgültige Diagnose", wurden somit in Abhängigkeit voneinander in den folgenden zwei Tabellen für Film 1 und Film 2 dargestellt.

Endgültige Diagnose	Beurteiler in %
ICD-Nr. 295.1 Hebephrenie	3 %
ICD-Nr. 295.8 Coenästhetische Schizophrenie	13 %
ICD-Nr. 296.2 Endogene Depression	19 %
ICD-Nr. 298.0 Reaktive psychot. Depression	3 %
ICD-Nr. 300.1 Hysterische Neurose	10 %
ICD-Nr. 300.4 Depressive Neurose	19 %
ICD-Nr. 300.6 Neurotisches Depersonalisations-Syndrom	4 %
ICD-Nr. 300.7 Hypochondrische Neurose	29 %

Abb. Nr. 3: Film 1. Endgültige Diagnosen nach dem 11. Informationsabschnitt als vierstellige, nicht in Hauptgruppen zusammengefaßte ICD-Nummern sämtlicher Beurteiler in Prozent.

Endgültige Diagnose	Beurteiler in %
ICD-Nr. 295.3 Paranoide Schizophrenie	37 %
ICD-Nr. 295.6 Residual-Schizophrenie	17 %
ICD-Nr. 295.7 Schizoaffektive Psychose	14 %
ICD-Nr. 296.1 Manie	11 %
ICD-Nr. 296.2 Endogene Depression	12 %
ICD-Nr. 300.3 Zwangsneurose	3 %
ICD-Nr. 300.4 Depressive Neurose	3 %
ICD-Nr. 316.0 Psychiatrisch unklare Fälle	3 %

Abb. Nr. 4: Film 2. Endgültige Diagnosen nach dem 9. Informationsabschnitt als vierstellige, nicht in Hauptgruppen zusammengefaßte ICD-Nummern sämtlicher Beurteiler in Prozent.

Nachdem jetzt die jeweiligen diagnostischen Entscheidungen aller Beurteiler zusammen in Prozent während der diagnostischen Interviews bzw. am Schluß der Übung demonstriert wurden, soll nun näher darauf eingegangen werden, ob die einzelnen Beurteiler ihre Entscheidungen von Anfang an beibehielten bzw. am Ende des Interviews änderten.

Hierfür wurden nun die vierstelligen ICD-Nummern, die von den Beurteilern mit dem höchsten subjektiven Sicherheitsgrad diagnostiziert worden waren, jeweils nach dem 1. Informationsanteil und nach dem 11. Informationsabschnitt für Film 1 in drei diagnostische Gruppen zusammengefaßt:

1. Neurotische Erkrankungen (ICD-Nr. 300 bis 305),
2. Endogene bzw. reaktive psychotische Depression (ICD-Nr. 296.2 bzw. 298.0),
3. Schizophrene Erkrankungen (ICD-Nr. 295).

In der folgenden Tabelle (Abb. Nr. 5) sind nun die endgültigen Diagnosen nach dem 11. Informationsabschnitt des 1. Filmes in Abhängigkeit von den "wahrscheinlichsten" Diagnosen mit dem höchsten subjektiven Sicherheitsgrad nach dem 1. Informationsabschnitt dargestellt. Die Prozentangaben beziehen sich jeweils auf die Anzahl der Beurteiler, die diese Diagnose mit dem höchsten subjektiven Sicherheitsgrad in den entsprechenden Informationsabschnitten (IA) angegeben hatten.

	Neurose 11. IA	Depression, endogen oder reakt. psychot. 11. IA	Schizophrenie 11. IA
Neurose 1. IA	<u>56 %</u>	10 %	-
Depression, endogen oder reakt. psychot. 1. IA	3 %	<u>14 %</u>	-
Schizophrenie 1. IA	17 %	-	-

Abb. Nr. 5: Film 1. Beziehung zwischen Anfangs- und Enddiagnose

Anahnd der obigen Darstellung wird ersichtlich, daß 70 % aller Beurteiler von Film 1 am Ende der diagnostischen Übung die gleiche endgültige Entscheidung, wie sie von ihnen bereits am Anfang der Untersuchung getroffen worden war, beibehielten. Nur insgesamt 30 % aller Beurteiler von Film 1 änderten während ihres diagnostischen Prozesses von Anfang bis Ende des Patienteninterviews ihre "wahrscheinlichste" Diagnose.

56 % aller Beurteiler, die am Anfang der Untersuchung als "wahrscheinlichste" Diagnose eine neurotische Erkrankung diagnostiziert hatten, behielten diese Entscheidung als endgültige Diagnose bis zum Schluß der Untersuchung bei bzw. kamen dann wieder darauf zurück. 14 % aller Beurteiler, die anfangs eine endogene bzw. reaktive psychotische Depression für die wahrscheinlichste Diagnose hielten, zeigten hinsichtlich ihrer endgültigen Diagnose keine Änderung. 10 % aller Beurteiler, die anfangs eine psychotische Depression für die wahrscheinlichste Diagnose hielten, änderten ihre endgültige Diagnose dahingehend, daß sie am Schluß der diagnostischen Übung eine Neurose für die endgültige Diagnose hielten, usw. .

Folgende Gruppen wurden wie oben bei Film 1 beschrieben nun auch für den 2. Film zusammengefaßt:

1. Schizophrene Erkrankungen (ICD-Nr. 295, mit Ausnahme von ICD-Nr. 295.7),
2. Schizoaffektive Psychose (ICD-Nr. 295.7),
3. Neurotische Erkrankungen (ICD-Nr. 300 bis 305),
4. Manisch-depressive Psychose (ICD-Nr. 296),
5. Körperlich begründbare psychiatrische Krankheitsbilder (ICD-Nr. 292, 293, 294, 309),
6. Psychiatrisch unklare Fälle (ICD-Nr. 316.0).

In der nachfolgenden Tabelle (Abb. Nr. 6) sind die endgültigen Diagnosen nach dem 9. Informationsabschnitt der Beurteiler von Film 2 in Prozent in Abhängigkeit von den "wahrscheinlichsten" Diagnosen nach dem 1. Informationsabschnitt dargestellt.

	ICD-Nr. 295 9. IA	ICD-Nr. 295.7 9. IA	ICD-Nr. 300-305 9. IA	ICD-Nr. 296 9. IA	ICD-Nr. 292, 293, 294, 309 9. IA	ICD-Nr. 316.0 9. IA
ICD-Nr. 295 1. IA	<u>34 %</u>	3 %	14 %	-	-	3 %
ICD-Nr. 295.7 1. IA	9 %	<u>3 %</u>	3 %	-	-	-
ICD-Nr. 300-305 1. IA	-	-	<u>11 %</u>	-	-	-
ICD-Nr. 296 1. IA	3 %	3 %	6 %	<u>3 %</u>		
ICD-Nr. 292, 293, 294, 309 1. IA	-	-	-	-	<u>3 %</u>	-
ICD-Nr. 316.0 1. IA	-	-	-	-	-	<u>3 %</u>

Abb. Nr. 6: Film 2. Beziehung zwischen Anfangs- und Enddiagnose

Es zeigte sich, daß bei 57 % aller Beurteiler von Film 2 die anfängliche diagnostische Entscheidung, also die Diagnose mit dem höchsten subjektiven Sicherheitsgrad, nach dem 1. Informationsabschnitt identisch war mit der endgültigen Diagnose nach dem 9. Informationsabschnitt. 43 % aller Beurteiler des 2. diagnostischen Interviews hatten ihre anfängliche diagnostische Entscheidung im Verlaufe des diagnostischen Prozesses geändert und eine andere ICD-Nr. als endgültige Diagnose mit dem höchsten subjektiven Sicherheitsgrad gewählt. Die jeweils unterstrichenen Prozentangaben in der Tabelle (Abb. Nr. 6) zeigen im Detail die Anzahl aller derjenigen Beurteiler an, deren anfängliche diagnosti-

sche Entscheidung mit ihrer endgültigen Diagnose übereinstimmt.

b) Klinisch-psychiatrische Erfahrung und Wahl der Diagnose

Alle Beurteiler waren in je zwei Gruppen eingeteilt worden: in eine Gruppe mit mehr als 2 1/2 Jahren klinischer Erfahrung in der Psychiatrie und in eine zweite Gruppe mit weniger als 2 1/2 Jahren klinisch-psychiatrischer Erfahrung.

Nach dem "Fisher's exact probability test" (Vierfeldertest) wurde nun berechnet, ob eine Korrelation der klinischen Erfahrung in der Psychiatrie der einzelnen Beurteiler einerseits und deren endgültigen Diagnosen andererseits besteht.

Für Film 2 konnte diesbezüglich ein statistisch signifikanter Unterschied festgestellt werden: diejenigen Beurteiler, die über weniger als 2 1/2 Jahre klinisch-psychiatrischer Erfahrung verfügten, diagnostizierten eher eine affektive Erkrankung als endgültige Diagnose nach dem 9. Informationsabschnitt als diejenigen Beurteiler, die über mehr als 2 1/2 Jahre klinischer Erfahrung in der Psychiatrie verfügten ($p = 0,037$).

In Film 1 konnte nach dem "Fisher's exact probability test" lediglich eine Tendenz ($p = 0,069$) dahingehend festgestellt werden, daß die Beurteiler, die über mehr als 2 1/2 Jahre klinischer Erfahrung in der Psychiatrie verfügten, eher eine schizophrene Erkrankung endgültig diagnostizierten als ihre jüngeren Kollegen.

2. Die subjektive Sicherheit in Abhängigkeit von der Diagnose

Als nächstes wurden für Film 1 und für Film 2 die Mittelwerte sämtlicher Sicherheitsgrade (SG) der einzelnen Diagnosengruppen erstellt und der Verlauf dieser Diagnosenkurven mit der jeweiligen Standardabweichung in den folgenden Abbildungen (Abb. 11 bis Abb. Nr. 17) graphisch dargestellt. Diese graphischen Darstellungen vermitteln nun einen Eindruck von dem Mittelwertstrend bzw. Dispersionstrend der einzelnen diagnostischen Hauptgruppen. Pro diagnostischer Hauptgruppe ergab sich je ein Diagramm, für Film 1 wurden also fünf Diagramme hergestellt, für Film 2 insgesamt sechs Diagramme.

SG% = Sicherheitsgrad in Prozent
IA = Informationsabschnitt
M = Mittelwert
S = Standardabweichung

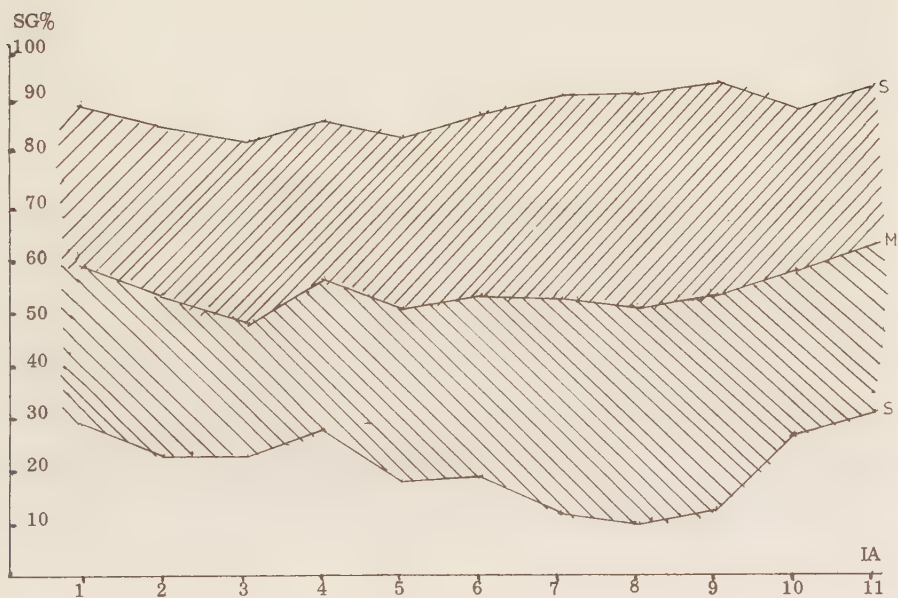


Abb. Nr. 7: Film 1

Mittlerer Sicherheitsgrad (± 1 Standardabweichung) der Psychiater, die eine Neurose bzw. Persönlichkeitsstörung diagnostizierten über die 11 Informationsanteile.

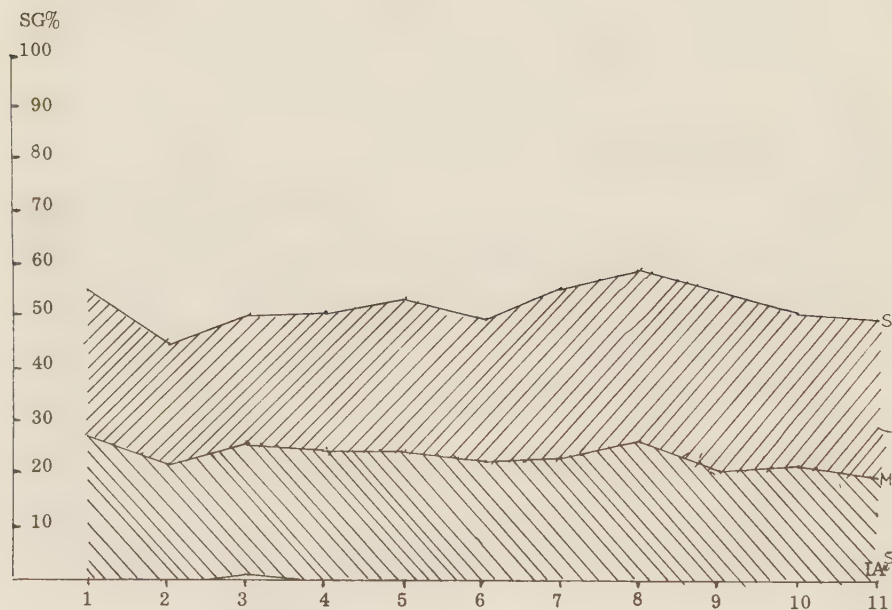


Abb. Nr. 8: Film 1

Mittlerer Sicherheitsgrad (± 1 Standardabweichung) der Psychiater, die eine manisch-depressive Psychose diagnostizierten über die 11 Informationsanteile.

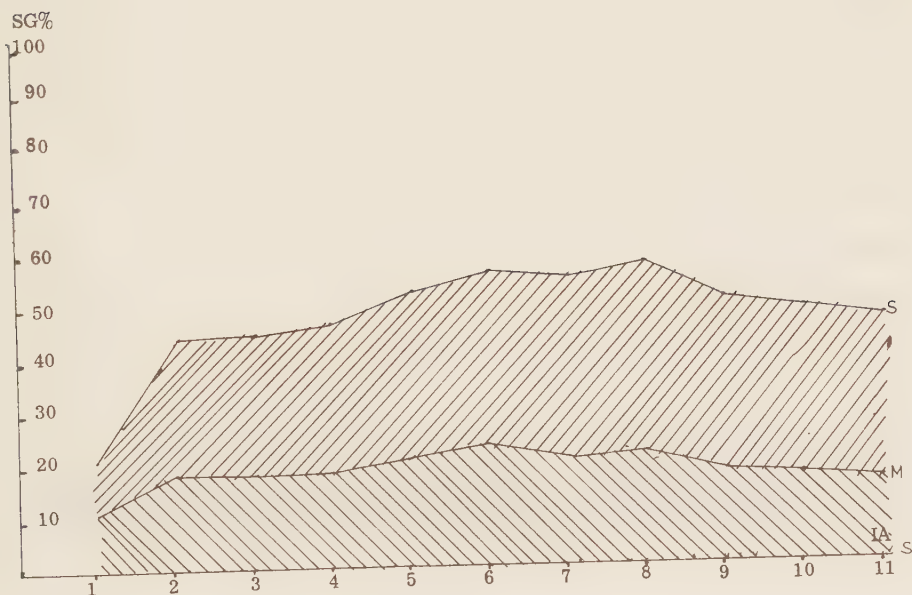


Abb. Nr. 9: Film 1

Mittlerer Sicherheitsgrad (± 1 Standardabweichung)
der Psychiater, die eine schizophrene Erkrankung
diagnostizierten über die 11 Informationsanteile.

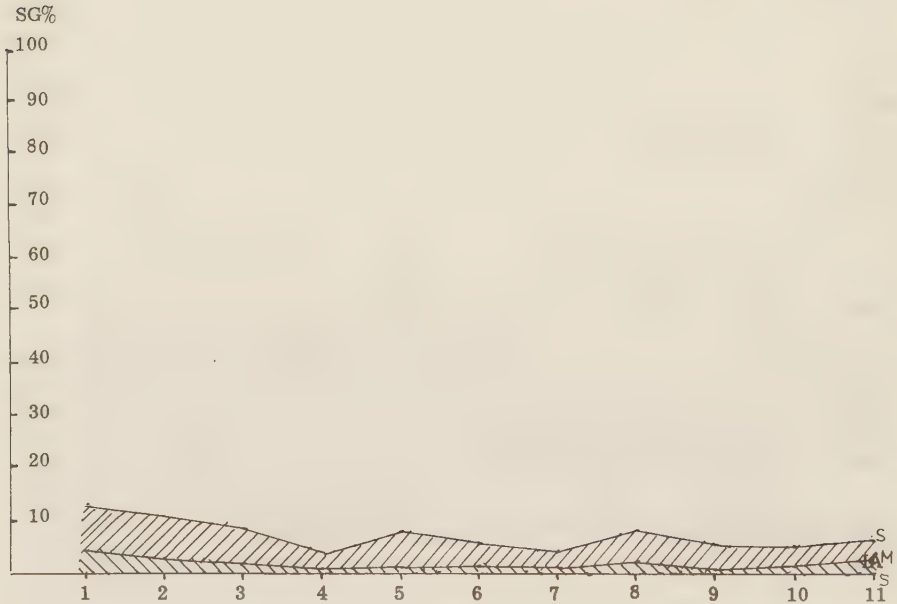


Abb. Nr. 10: Film 1

Mittlerer Sicherheitsgrad ($\pm$ Standardabweichung) der Psychiater, die eine körperlich begründbare psychiatrische Erkrankung diagnostizierten über die 11 Informationsanteile.



Abb. Nr. 11: Film 1

Mittlerer Sicherheitsgrad (± 1 Standardabweichung)
der Psychiater, die die Diagnose stellten "psychiatrisch
unklare Fälle" (ICD-Nr. 316.0) über die 11 Informa-
tionsanteile.

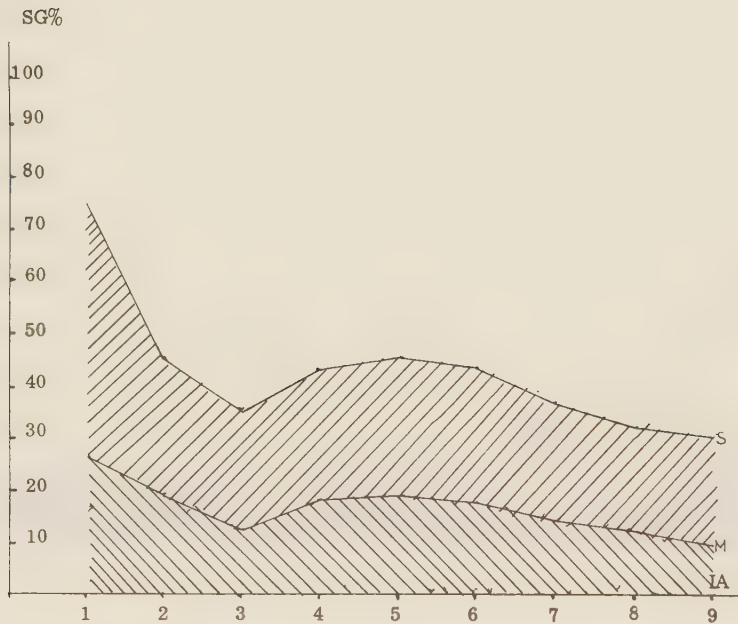


Abb. Nr. 12: Film 2

Mittlerer Sicherheitsgrad (± 1 Standardabweichung)
der Psychiater, die eine Neurose bzw. Persönlich-
keitsstörung diagnostizierten über die 9 Informations-
anteile.

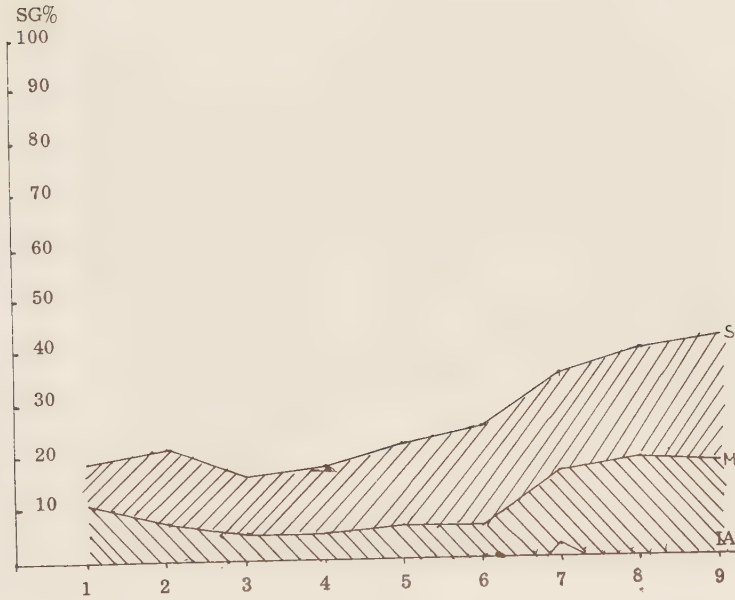


Abb. Nr. 13: Film 2

Mittlerer Sicherheitsgrad (± 1 Standardabweichung)
der Psychiater, die eine manisch-depressive Psychose
bzw. eine reaktive depressive Psychose diagnostizierten
über die 9 Informationsanteile.

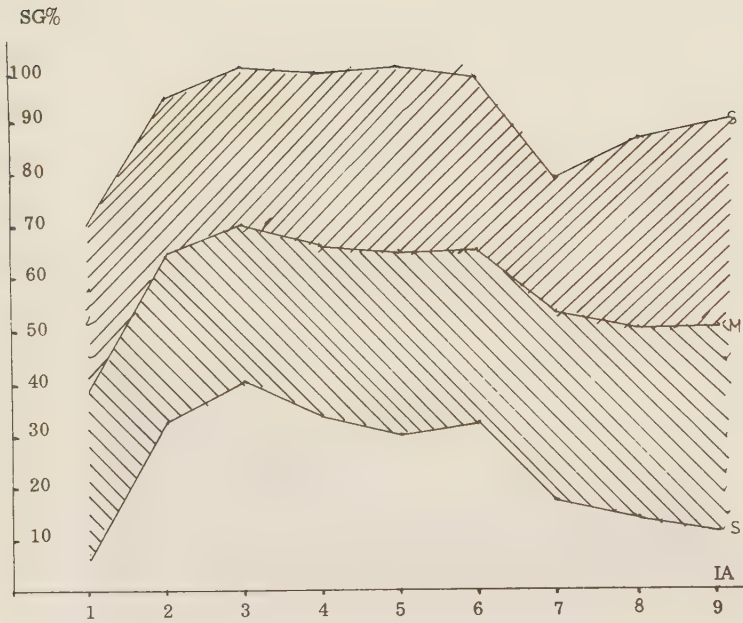


Abb. Nr. 14: Film 2

Mittlerer Sicherheitsgrad (± 1 Standardabweichung)
der Psychiater, die eine schizophrene Erkrankung
diagnostizierten über die 9 Informationsanteile.



Abb. Nr. 15: Film 2

Mittlerer Sicherheitsgrad (± 1 Standardabweichung)
der Psychiater, die eine körperlich begründbare
psychiatrische Erkrankung diagnostizierten über die
9 Informationsanteile.

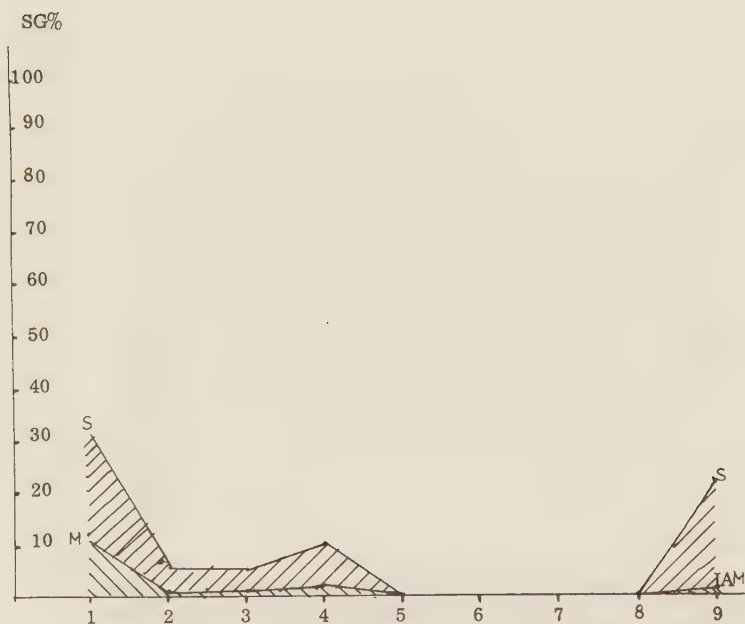


Abb. Nr. 16: Film 2

Mittlerer Sicherheitsgrad (± 1 Standardabweichung) der Psychiater, die als Diagnose "psychiatrisch unklare Fälle" (ICD-Nr. 316.0) stellten über die 9 Informationsanteile.

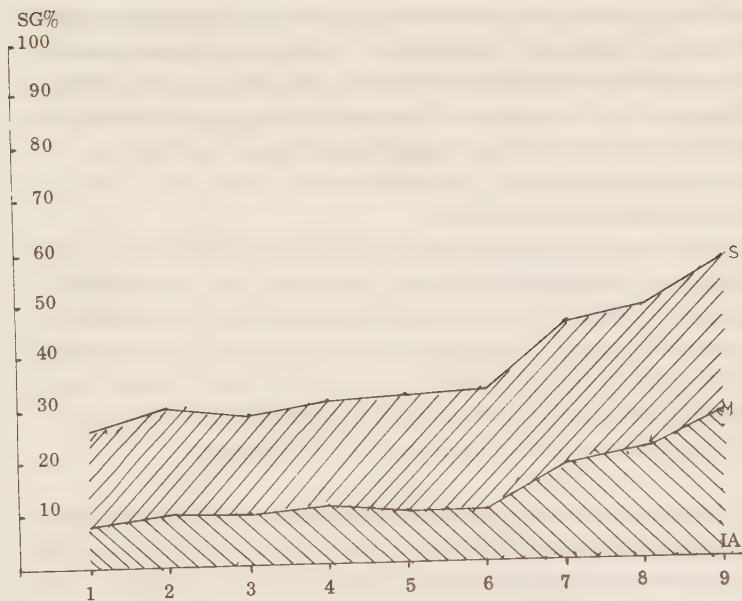


Abb. Nr. 17: Film 2

Mittlerer Sicherheitsgrad (± 1 Standardabweichung) der Psychiater, die eine Schizoaffektive Psychose (ICD-Nr. 295.7) diagnostizierten über die 9 Informationsanteile.

Es zeigt sich anhand dieser graphischen Darstellungen ein relativ gleichmäßiger Verlauf der Mittelwerte eigentlich aller diagnostischer Hauptgruppen: Der Sicherheitsgrad nimmt global gesehen für alle Psychiater durch die in den Fernsehaufzeichnungen gebotenen Informationsabschnitte hindurch weder zu, noch nimmt dieser ab, d. h. daß sich insgesamt immer etwa gleich viele Beurteiler für eine bestimmte Diagnose entscheiden. Diese Aussage gilt im großen und ganzen für alle diagnostischen Hauptgruppen, wobei eine stärkere Abweichung von den übrigen Darstellungen vielleicht der Verlauf der Mittelwerte der 5. diagnostischen Hauptgruppe (ICD-Nr. 316.0 "psychiatrisch unklare Fälle", Abb. Nr. 16) aufweist. Letztgenannte Diagnose wurde jedoch nur von einer äußerst geringen Anzahl der Beurteiler in Erwägung gezogen. Der Kurvenverlauf für die Mittelwerte der übrigen diagnostischen Hauptgruppen kann im wesentlichen also als 'horizontal' bezeichnet werden, was für Film 1 noch deutlicher erkennbar ist als für Film 2.

Die dazugehörige jeweilige Standardabweichung drückt schließlich durch den gleichfalls im wesentlichen horizontalen Verlauf aus, daß der subjektive Sicherheitsgrad in Prozent (SG%), wiederum global gesehen für alle Diagnostiker, über sämtliche Filmunterbrechungen hindurch hinsichtlich der jeweiligen diagnostischen Hauptgruppen sich relativ gleichförmig verhält. Dies gilt ebenfalls wieder für beide diagnostische Interviews, wobei auch hier wieder in Film 1 der beschriebene Kurvenverlauf deutlicher erkennbar ist.

Anstelle der eben geschilderten Mittelwertsberechnung mit der jeweiligen Standardabweichung für die subjektiven Sicherheiten sämtlicher Beurteiler und aller diagnostischer Hauptgruppen wurden schließlich auch die sog. Medianwerte für die entsprechenden Daten berechnet und graphisch dargestellt. Die Medianwert-Darstellungen weisen jedoch keine bemerkenswerten Unterschiedlichkeiten bezüglich der entsprechenden Mittelwertsergebnisse auf. Es wurde deshalb auf eine Abbildung der graphischen Darstellungen der Medianwerte verzichtet. Auch sie zeigen im wesentlichen einen horizontalen Verlauf. Dies unterstützt die aus den Mittelwertsergebnissen gewonnene Aussage, nämlich daß die subjektiven Sicherheiten sämtlicher Beurteiler bezüglich aller diagnostischen Hauptgruppen durch alle Informationsabschnitte beider Fernsehaufzeichnungen hindurch im wesentlichen konstant bleiben.

Dieselben Mittelwertskurven, die oben abgebildet sind, werden in den nachfolgenden zwei Tabellen in Zahlen ausgedrückt nochmals wiedergegeben (Abb.

Nr. 18 und 19). Die Zahlenbilder veranschaulichen noch einmal, daß die von den Beurteilern angegebenen subjektiven Sicherheiten in Prozent (SG%) einen verhältnismäßig konstant bleibenden mittleren Sicherheitsgrad aller Beurteiler zusammen aufweisen. Nennenswerte Veränderungen dieser Zahlenbilder im Verlaufe der diagnostischen Informationsabschnitte (IA) lassen sich wohl kaum erkennen. Es gilt also wieder die Aussage, daß global gesehen für alle Diagnostiker stets eine konstante subjektive Sicherheit bezüglich ihrer vorgeschlagenen Diagnosen besteht, d. h. daß die Sicherheit am Ende der Untersuchung nicht größer ist als nach den ersten 4 bzw. 4 1/2 Minuten der Interviews. Lediglich in einzelnen diagnostischen Hauptgruppen bestehen gewisse Abweichungen von diesem geschilderten Trend.

	1. IA	2. IA	3. IA	4. IA	5. IA	6. IA	7. IA	8. IA	9. IA	10. IA	11. IA
1. H. Gr. SG%	59 %	54 %	52 %	58 %	51 %	54 %	53 %	51 %	54 %	58 %	63 %
2. H. Gr. SG%	27 %	22 %	26 %	25 %	25 %	23 %	24 %	26 %	27 %	22 %	20 %
3. H. Gr. SG%	6 %	18 %	18 %	18 %	21 %	23 %	21 %	22 %	18 %	17 %	16 %
4. H. Gr. SG%	4 %	3 %	1 %	-	1 %	1 %	1 %	1 %	1 %	1 %	1 %
5. H. Gr. SG%	4 %	2 %	2 %	2 %	2 %	2 %	2 %	2 %	1 %	2 %	-

Abb. Nr. 18: Film 1

Mittelwerte der subjektiven Sicherheitsgrade (SG%) sämtlicher Beurteiler für alle fünf diagnostischen Hauptgruppen während der 11 Informationsabschnitte.

	1. IA	2. IA	3. IA	4. IA	5. IA	6. IA	7. IA	8. IA	9. IA
1. H. Gr. SG%	36 %	19 %	13 %	18 %	18 %	17 %	13 %	12 %	9 %
2. H. Gr. SG%	6 %	7 %	5 %	5 %	6 %	6 %	16 %	19 %	18 %
3. H. Gr. SG%	38 %	63 %	70 %	66 %	64 %	65 %	53 %	50 %	50 %
4. H. Gr. SG%	4 %	1 %	1 %	1 %	2 %	2 %	2 %	1 %	1 %
5. H. Gr. SG%	6 %	-	-	2 %	-	-	-	-	4 %
6. H. Gr. SG%	8 %	10 %	10 %	11 %	9 %	10 %	18 %	21 %	18 %

Abb. Nr. 19: Film 2

Mittelwerte der subjektiven Sicherheitsgrade (SG%) sämtlicher Beurteiler für alle sechs diagnostischen Hauptgruppen während der 9 Informationsabschnitte.

3. Die Persönlichkeit des Psychiaters

Die Halbform des Freiburger Persönlichkeitsinventars (FPI-A) war von sämtlichen Ärzten unmittelbar nach Beendigung der beiden diagnostischen Übungen ausgefüllt worden. Die Interpretation der individuellen Testergebnisse erfolgte nach der "Handanweisung für die Durchführung und Auswertung des Freiburger Persönlichkeitsinventares" von FAHRENBURG und SELG (1970). Das Profil der Mittelwerte aller FPI-Standardskalen ist für sämtliche Beurteiler zusammen auf einem FPI-Auswertungsbogen (Abb. Nr. 20) wiedergegeben. Es zeigen sich anhand dieses Profiles keine nennenswerten Abweichungen von den mittleren Normwerten der FPI-Standardskalen. Die Gesamtheit der an der Untersuchung beteiligten Psychiater kann demnach bezüglich ihrer Persönlichkeitseigenschaften nach dem Freiburger Persönlichkeitsinventar als "durchschnittlich" und "normal" eingestuft werden.

Abb. Nr. 20: Mittlere FPI-Profile
aller Beurteiler.

FPI

Gesamtform
Halbform A-B

Datum

Auswertungsbogen

		Prozent	4	7	12	17	20	17	12	7	4		
Skala	Rohwert	Standardwert	9	8	7	6	5	4	3	2	1	Stanine	
FPI 1		Nervosität psychosomatisch gestört	•	•	•	•	•	•	•	•	•		psychosomat. nicht gestört
FPI 2		Aggressivität aggressiv, emotional unreif	•	•	•	•	•	•	•	•	•		nicht aggressiv, beherrscht
FPI 3		Depressivität mißgestimmt, selbstunsicher	•	•	•	•	•	•	•	•	•		zufrieden, selbstsicher
FPI 4		Erregbarkeit reizbar, leicht frustriert	•	•	•	•	•	•	•	•	•		ruhig, stumpf
FPI 5		Geselligkeit gesellig, lebhaft	•	•	•	•	•	•	•	•	•		ungesellig, zurückhaltend
FPI 6		Gelassenheit selbstvertrauend, gutgelaunt	•	•	•	•	•	•	•	•	•		irritierbar, zögernd
FPI 7		Dominanzstreben sich durchsetzend, streng	•	•	•	•	•	•	•	•	•		nachgiebig, gemäßigt
FPI 8		Gehemmtheit gehemmt, gespannt	•	•	•	•	•	•	•	•	•		ungezwungen, kontaktfähig
FPI 9		Offenheit offen, selbstkritisch	•	•	•	•	•	•	•	•	•		verschlossen, unkritisch
FPI E		Extraversion extravertiert	•	•	•	•	•	•	•	•	•		introvertiert
FPI N		Emot. Labilität emotional labil	•	•	•	•	•	•	•	•	•		emotional stabil
FPI M		Maskulinität typisch männliche Selbstschilderung	•	•	•	•	•	•	•	•	•		typisch weibl. Selbstschildg.

54 %

54 %

Die Mittelwerte der einzelnen FPI-Standardskalen von allen Psychiatern, die im 1. Film eine Psychose (P) als endgültige Diagnose diagnostizierten, wurden verglichen mit denen jener Psychiater, die im 1. Film eine Neurose diagnostizierten. Mit Hilfe des t-Test-Verfahrens (t-Test unverbunden aus Rohdaten) für den Vergleich zweier Mittelwerte unabhängiger Stichproben wurde berechnet, ob statistisch signifikante Unterschiede hinsichtlich der Persönlichkeitseigenschaften nach dem Freiburger Persönlichkeitsinventar dieser beiden Beurteilergruppen bestehen.

Es konnten statistisch signifikante Unterschiede der FPI-Standardskalen auf dem 0,05-Niveau bezüglich des Merkmals "Aggressivität" festgestellt werden: die Diagnostikergruppe P, die als endgültige Diagnose irgendeine Form von Psychose diagnostiziert hatte, zeigte nach dem Freiburger Persönlichkeitsinventar eine statistisch signifikante "höhere" Aggressivität als die Diagnostikergruppe N, die als endgültige Diagnose irgendeine neurotische Erkrankung diagnostiziert hatte. Bei der Diagnostikergruppe N zeigte sich demnach umgekehrt eine statistisch signifikante "niedrigere" Aggressivität als bei der Diagnostikergruppe P ($p < 0,05$).

Für das Persönlichkeitsmerkmal "Extraversion" ließ sich eine Tendenz dahingehend erkennen, daß die Diagnostikergruppe P eine "größere" Extraversion zeigte als die Diagnostikergruppe N von Film 1 ($p < 0,1$). Die Diagnostikergruppe N dagegen zeigte eine "geringere" Extraversion nach dem Freiburger Persönlichkeitsinventar. Beurteiler des 1. Filmes, die eine Neurose als endgültige Diagnose stellten, zeigten sich also eher introvertiert, die Beurteiler, die eine Psychose diagnostiziert hatten, demnach eher extravertiert.

Bei Film 2 wurde ein ähnlicher Vergleich durchgeführt. Es wurden die Mittelwerte der einzelnen FPI-Standardskalen von allen Psychiatern, die eine affektive Erkrankung (A) als endgültige Diagnose stellten, verglichen mit jenen aller Psychiater, die eine schizophrene Erkrankung (S) diagnostizierten. Mittels des t-Test-Verfahrens konnte lediglich eine Tendenz ($p < 0,1$) der Diagnostikergruppen A und S hinsichtlich der Persönlichkeitseigenschaft "Erregbarkeit" nach dem Freiburger Persönlichkeitsinventar wie folgendermaßen festgestellt werden: die Diagnostikergruppe A, die als endgültige Diagnose eine affektive Erkrankung diagnostiziert hatte, neigte eher dazu, eine "niedrigere" Erregbarkeit aufzuweisen als die Diagnostikergruppe S, die als endgültige Diagnose eine schizophrene Erkrankung diagnostiziert hatte. Entsprechend neigte die Diagnostikergruppe S eher zu einer "höheren" Erregbarkeit als die Diagnostikergruppe A.

Weitere Aussagen über die Korrelation von Persönlichkeitsstörungen nach dem Freiburger Persönlichkeitsinventar mit den endgültigen diagnostischen Entscheidungen der jeweiligen Beurteiler ließen sich mittels des t-Test-Verfahrens allerdings für keine der beiden Fernsehaufzeichnungen machen.

4. Die Symptombeobachtung in Abhängigkeit von der Wahl der Diagnose

Als nächstes wurde nun versucht, eine Beziehung herzustellen zwischen der Wahl der Diagnose am Schluß der diagnostischen Interviews und den beobachteten Symptomen nach den Belegblättern des AMP-Systems (3. Psych. 3), die die Ärzte zur Beurteilung des psychischen Befundes der Patienten nach jeder Filmunterbrechung ausgefüllt hatten.

In Film 1 fielen pro Beurteiler 11 AMP-Belegblätter an, in Film 2 waren es 9 AMP-Belegblätter.

Die einzelnen Items des AMP-Belegblattes 3 für den psychischen Befund wurden für diese Untersuchung in bestimmte Merkmalsgruppen zusammengefaßt:

1. Bewußtseinsstörungen,
2. Orientierungsstörungen,
3. Aufmerksamkeits- und Gedächtnisstörungen,
4. Formale Denkstörungen,
5. Überwertige Ideen, Zwänge, Phobien,
6. Wahn,
7. Sinnestäuschungen
8. Ich- und Persönlichkeitsstörungen,
9. Verstimmungen, Gefühlsstörungen,
10. Psychomotorische Störungen,
11. Verlaufsbesonderheiten,
12. Störungen des Trieb- und Sozialverhaltens, Pflegebedürftigkeit.

Alle Items einer bestimmten Merkmalsgruppe wurden für jeden einzelnen Beurteiler über alle Informationsabschnitte aufgelistet und summiert, wobei die sich wiederholenden Merkmale jeweils nur einfach gezählt wurden. Hatte beispielsweise ein bestimmter Beurteiler das Merkmal "klagsam/jammerig" bei jedem Informationsabschnitt, also 11mal in Film 1 und 9mal in Film 2, als vorhanden ("leicht", "mittel" oder "stark") angegeben, so war dies insgesamt

nur als ein beobachtetes Symptom, zugehörig zur 9. Merkmalsgruppe ("Verstimmungen, Gefühlsstörungen") bewertet worden. Die auf diese Weise erhaltenen Einzelsummen aller Items der genannten Merkmalsgruppen wurden ganz am Schluß zu der sog. "totalen Summe" summiert, durch die sich erkennen läßt, wieviele verschiedene Items insgesamt die einzelnen Beurteiler während einer diagnostischen Übung beobachtet hatten.

Beispiel: Ein Beurteiler hatte aus den 12 einzelnen Merkmalsgruppen in ihrer oben beschriebenen Reihenfolge je 1-0-3-2-2-1-1-1-2-1-11-19 verschiedene Items während eines diagnostischen Interviews beobachtet und auf seinen Belegblättern vermerkt. Es ergab sich daraus für diesen Beurteiler eine "totale Summe" von insgesamt 44 verschiedenen von ihm wahrgenommenen Merkmalen während dieser Untersuchung.

Die Mittelwerte der einzelnen Merkmalssummen und der "totalen Summen" von allen Psychiatern, die im ersten Film eine Psychose (P) als endgültige Diagnose diagnostizierten, wurden in einem weiteren Schritt nun verglichen mit denen aller jener Psychiater, die im ersten Film eine Neurose (N) diagnostizierten. Mittels des t-Test-Verfahrens konnte schließlich festgestellt werden, daß keine statistisch signifikanten Unterschiede zwischen diesen beiden Diagnostikerguppen bezüglich der Häufigkeit festgestellter bzw. beobachteter Symptome nach dem AMP-System bestehen.

Gleichermaßen wurden die Mittelwerte der einzelnen Merkmalssummen und der "totalen Summe" von allen jenen Beurteilern, die im 2. Film eine affektive Erkrankung (A) endgültig diagnostiziert hatten, verglichen mit denen der Beurteiler, die eine schizophrene Erkrankung (S) diagnostiziert hatten. Auch hier konnten keine statistisch signifikanten Unterschiede hinsichtlich der Art und der Anzahl beobachteter Symptome festgestellt werden.

Für Film 1 ergab sich eine Ausnahme: Unter der 11. Merkmalsgruppe "Verlaufsbesonderheiten" zeigten sich statistisch signifikante Unterschiede ($p < 0,05$) dahingehend, daß die Beurteiler, die eine Psychose (P) diagnostizierten, häufiger psychopathologische Symptome in Form von Items nach dem AMP-System beobachteten, als die Beurteiler, die eine Neurose (N) diagnostizierten.

Verlaufsbesonderheiten	Anzahl der Beurteiler N (Neurose)	Anzahl der Beurteiler P (Psychose)
"abends besser"	5	25
"morgens besser"	-	-
"nächtliche Exazerbationen"	-	-
"andere Verlaufsbesonderheiten"	-	1
Items-Summe	5	26

Abb. Nr. 21: Film 1

Die Items der Merkmalsgruppe "Verlaufsbesonderheiten" wie sie von den beiden Psychiatergruppen P und N beobachtet wurden.

Die Ergebnisse des t-Tests lassen also folgende allgemeine Aussage zu: die Wahl der Diagnose der Ärzte hatte im allgemeinen keinen Einfluß auf die Symptomwahrnehmung bzw. die Anzahl und die Art der beobachteten Items nach dem AMP-System, umgekehrt übte diese Symptomwahrnehmung offenbar keinen meßbaren Einfluß auf die endgültige Diagnose aus. Dies gilt mit der einzigen, oben erläuterten Ausnahme sowohl für das erste als auch für das zweite diagnostische Interview.

5. Die Symptombeobachtung in Abhängigkeit von der klinisch-psychiatrischen Erfahrung der Beurteiler

Es gilt wieder die Einteilung der Psychiater in zwei Gruppen: in eine Gruppe mit mehr als 2 1/2 Jahren klinisch-psychiatrischer Erfahrung und in eine zweite Gruppe mit weniger als 2 1/2 Jahren klinisch-psychiatrischer Erfahrung.

Die Mittelwerte der Itemssummen aller jener Psychiater mit mehr als 2 1/2 Jahren klinischer Erfahrung wurden mittels des t-Test-Verfahrens verglichen mit den Mittelwerten der Itemssummen aller Psychiater mit weniger als 2 1/2 Jahren klinischer Erfahrung.

Es ergaben sich statistisch signifikante Unterschiede ($p < 0,05$) bei Film 1 für vier verschiedene Merkmalsgruppen: "Wahn", "Sinnestäuschungen", "Verstimmungen, Gefühlsstörungen" und "Störungen des Trieb- und Sozialverhaltens". Die weniger erfahrenen Ärzte hatten jeweils eine statistisch signifikante geringere Anzahl verschiedener Items dieser genannten Merkmalsgruppen beobachtet, als deren erfahreneren Kollegen.

In derselben Richtung ($p < 0,1$) zeigte sich auch eine Tendenz hinsichtlich der Merkmalsgruppe "Überwertige Ideen, Zwänge, Phobien" und schließlich auch bezüglich der "totalen Summe" sämtlicher Items aus allen Merkmalsgruppen zusammen.

Die erfahreneren Psychiater tendierten demnach offenbar mehr dazu, häufiger Symptome als "vorhanden" zu betrachten, als ihre weniger erfahrenen Kollegen.

Entsprechende Resultate ergaben sich auch bei Film 2. Hier zeigten sich statistisch signifikante Unterschiede ($p < 0,05$) der Items-Summen zweier Merkmalsgruppen ("Überwertige Ideen, Zwänge, Phobien"; "Verlaufsbesonderheiten") einerseits und der klinischen Erfahrung der Psychiater andererseits.

Kliniker mit weniger als 2 1/2 Jahren Erfahrung in der Psychiatrie gaben mit statistischer Signifikanz weniger Symptome in den AMP-Belegblättern an, Kliniker mit mehr als 2 1/2 Jahren Erfahrung in der Psychiatrie gaben mit statistischer Signifikanz entsprechend mehr Items als "vorhandene Symptomatik" an. Eine derartige Tendenz ($p < 0,1$) zeigte sich bei Film 2 schließlich auch bezüglich der Merkmalsgruppe "Aufmerksamkeits- und Gedächtnisstörungen". Auch hier war die Anzahl der beobachteten psychopathologischen Symptome bei den erfahreneren Klinikern höher als bei ihren unerfahreneren Kollegen.

In den folgenden beiden Tabellen sind diese geschilderten Ergebnisse nochmals zahlenmäßig angeführt. Die Zahlen drücken aus, wieviele Items insgesamt von den jeweiligen Beurteilern (mehr bzw. weniger als 2 1/2 Jahre klinische Erfahrung in der Psychiatrie) beobachtet wurden.

Merkmalgruppen	Psychiater mit weniger als 2 1/2 Jahren klin. Erfahrung (n = 14)	Psychiater mit mehr als 2 1/2 Jahren klin. Erfahrung (n = 17)
"Wahn"	9	23
"Sinnesstörungen"	13	21
"Verstimmungen, Gefühlsstörungen"	123	205
"Störungen des Trieb- und Sozialverhaltens"	27	39

Abb. Nr. 22: Film 1

Anzahl der Items aus einzelnen Merkmalsgruppen, wie sie von den erfahreneren und den weniger erfahrenen Beurteilern beobachtet wurden.

Merkmalgruppen	Psychiater mit weniger als 2 1/2 Jahren klin. Erfahrung (n = 16)	Psychiater mit mehr als 2 1/2 Jahren klin. Erfahrung (n = 19)
"Überwertige Ideen, Zwänge, Phobien"	25	39
"Verlaufsbesonder- heiten"	2	14
"Aufmerksamkeits- und Gedächtnis- störungen"	11	20

Abb. Nr. 23: Film 2

Anzahl der Items aus einzelnen Merkmalsgruppen, wie sie von den erfahreneren und den weniger erfahrenen Beurteilern beobachtet wurden.

6. Beschreibung des diagnostischen Prozesses

a) Einteilung und Interpretation der sog. "Einzeldiagramme"

Die im Anhang abgebildeten sog. "Einzeldiagramme" sollen den gesamten diagnostischen Prozess jedes einzelnen Beurteilers der Fernsehaufzeichnungen als Funktion der pro Informationsabschnitt (IA) von diesen Beurteilern in Erwägung gezogenen Diagnosen und Differentialdiagnosen in Abhängigkeit von dem jeweils dazugehörigen subjektiven Sicherheitsgrad in Prozent (SG%) graphisch darstellen.

Es wurde versucht, diese sog. "Einzeldiagramme" nach folgenden Merkmalen zu definieren und einzuteilen:

1. Ein Merkmal für diese Interpretation war die Anzahl der Kreuzungen und Berührungen der einzelnen Diagnosen-Verläufe bis zum Schluß der Interviews. Die Kreuzungen und Berührungen der Kurven auf der Nulllinie wurden grundsätzlich nicht gewertet. Die Benennung dieses Merkmals der Bewegungen und Wellen, die das Bild eines solchen Diagrammes prägen, bezeichneten wir schließlich wertfrei als "Flexibilität".
2. Eine weitere Richtlinie, diese "Einzeldiagramme" einzuteilen, war die Differenz der subjektiven Sicherheitsgrade (SG%) zwischen der Hauptdiagnose ("wahrscheinlichste Diagnose" nach jedem Informationsanteil bzw. "endgültige Diagnose" nach dem letzten Informationsanteil) und der Differentialdiagnose, die mit dem jeweils zweithöchsten subjektiven Sicherheitsgrad von dem entsprechenden Beurteiler diagnostiziert worden war. Diese Differenz wurde über alle Filmunterbrechungen hinweg summiert. Die Summen wurden schließlich - wieder völlig wertfrei gemeint - als sog. "Sicherheit" eines Beurteilers bezeichnet.
3. Das dritte Kriterium für die Einordnung dieser "Einzeldiagramme" in gesonderte Gruppen war schließlich die "Diagnosen-Summe", d. h. die Summe sämtlicher, auch wiederholt während eines diagnostischen Prozesses auftretenden Diagnosen bzw. diagnostischen Hauptgruppen (siehe oben!), die von jedem Beurteiler während der diagnostischen Interviews in Erwägung gezogen worden waren. Wäre z. B. ausschließlich die Diagnose einer Schizophrenie in allen 11 Informationsabschnitten des 1. Filmes genannt worden, so hätte dies eine "Diagnosensumme" von 11 ergeben etc., die kleinstmögliche "Diagnosen-Summe" von Film 1 war

also 11, die kleinstmögliche "Diagnosen-Summe" von Film 2 war entsprechend 9.

Die durchschnittliche "Flexibilität" (s. unter Punkt 1) lag beim 1. Film bei 2,3 Kreuzungen und Berührungen bzw. bei 3,0 Kreuzungen und Berührungen beim 2. Film.

Nun unterteilten wir die Beobachter des ersten und des zweiten Filmes in je zwei Gruppen, je nachdem, ob sie in ihrer "Flexibilität" unter oder über dem Durchschnittswert lagen. Es ergaben sich demnach eine Gruppe 1 mit "niedriger Flexibilität" und eine Gruppe 2 mit "hoher Flexibilität".

Die durchschnittliche "Sicherheit" betrug für Film 1 641 Gesamtdifferenz zwischen sämtlichen subjektiven Sicherheitsgraden (SG%) der Haupt- und Zweitdiagnosen aller Beurteiler, für Film 2 betrug diese durchschnittliche "Sicherheit" 491.

Auch hier unterteilten wir die Beurteiler wieder in je zwei Gruppen: in eine Gruppe 1 mit "hoher Sicherheit" und in eine Gruppe 2 mit "niedriger Sicherheit".

Der Mittelwert der "Diagnosen-Summe" von Film 1 betrug 21,3 Diagnosen pro Beurteiler, von Film 2 betrug er 18,4 Diagnosen. Beurteiler mit einer "niedrigen Diagnosen-Summe" wurden nun der Gruppe 1 zugeordnet, Beurteiler mit einer "hohen Diagnosen-Summe" wurden der Gruppe 2 zugeordnet.

<u>"Flexibilität"</u>		<u>"Sicherheit"</u>		<u>"Diagnosen-Summe"</u>	
niedrig	hoch	niedrig	hoch	niedrig	hoch
Gruppe 1	Gruppe 2	Gruppe 2	Gruppe 1	Gruppe 1	Gruppe 2

Abb. Nr. 24: "Einzeldiagramme": Einteilung in zwei Gruppen entsprechend bestimmten Merkmalen.

Eine "hohe Flexibilität" und eine "hohe Diagnosen-Summe" drücken ein gewisses diagnostisches Schwanken des jeweiligen Beurteilers aus. Sein spezifisches "Einzeldiagramm" zeigt viele Berührungen und Kreuzungen der einzelnen Diagnosenverläufe und drückt anhand der relativ hohen "Diagnosen-Summe" aus,

daß er viele differentialdiagnostische Überlegungen in Erwägung gezogen hatte.

Große, über dem Durchschnittswert liegende Differenzen zwischen den subjektiven Sicherheitsgraden (SG%) der Haupt- und Zweitdiagnosen eines Diagnostikers insgesamt drücken aus, daß sich der Beurteiler seiner jeweiligen Hauptdiagnosen verhältnismäßig sicher war ("hohe Sicherheit", Gruppe 2). Weniger sicher war sich der Beurteiler seiner jeweiligen Hauptdiagnosen, wenn die Differenz zwischen den subjektiven Sicherheitsgraden von Haupt- und Zweitdiagnosen unter dem Durchschnittswert lagen, also relativ klein waren.

Jeder einzelne Beurteiler wurde hinsichtlich dieser drei Charakteristika ("Flexibilität", "Sicherheit", "Diagnosen-Summe") beschrieben.

Gemäß diesen erläuterten Charakteristika wurden die Psychiater nochmals in zwei Übergruppen eingeteilt: je nachdem, ob er bei diesen drei Hauptmerkmalen vorwiegend der Gruppe 1 oder vorwiegend der Gruppe 2 angehörte.

Beispiel: Ein Beurteiler mit einer "hohen Flexibilität" im oben definierten Sinne, einer "hohen Sicherheit" und einer "niedrigen Diagnosen-Summe" erhielt die einzelnen Gruppenbewertungen 2-1-1 (vgl. Tabelle oben!), woraus sich gewissermaßen als Mittelwert die Gruppeneinteilung 1 ergab. Der Beurteiler zeigte zwar verhältnismäßig viele Schwankungen in seinen Überlegungen ("hohe Flexibilität"), war sich der jeweiligen Hauptdiagnose aber immer sehr sicher ("hohe Sicherheit") und zog auch relativ wenige Differentialdiagnosen in Erwägung ("niedrige Diagnosen-Summe").

Ein weiteres Beispiel: Ein Beurteiler mit einer "niedrigen Flexibilität", einer "hohen Sicherheit" und einer "niedrigen Diagnosen-Summe" erhielt die Gruppenbewertungen 1-1-1, woraus sich wieder die mittlere Gruppeneinteilung 1 ergab. Dieser Beurteiler zeigte verhältnismäßig wenige Schwankungen in seinen diagnostischen Überlegungen ("niedrige Flexibilität"), war sich seiner jeweiligen Hauptdiagnose pro Informationsabschnitt relativ sicher ("hohe Sicherheit") und dachte während seines diagnostischen Prozesses nur an wenige Differentialdiagnosen ("niedrige Diagnosen-Summe").

Abbildung Nr. 25 stellt nun ein graphisches Beispiel eines solchen "Einzeldiagrammes" aus Film 1 dar, das endgültig zur Gruppe 1 eingeteilt worden war. In diesem speziellen Beispiel verhielten sich die einzelnen Gruppenbewertungen folgendermaßen: nur eine sog. Berührung, keine Kreuzungen (Gruppe 1), eine "hohe Sicherheit" von 890 (Gruppe 1), eine "niedrige

Diagnosen-Summe" von insgesamt 14 erwähnten Diagnosen (Gruppe 1), woraus sich also die endgültige Zuordnung zur Übergruppe 1 ergab.

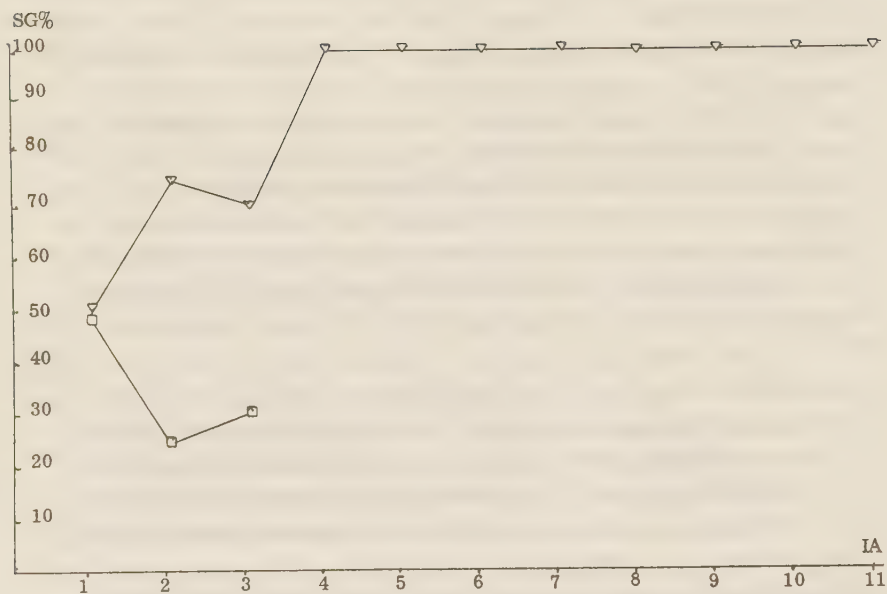
Der Beurteiler änderte seine diagnostischen Überlegungen während des gesamten diagnostischen Interviews kaum, ab dem 4. Informationsabschnitt war er sich seiner Hauptdiagnose vollkommen sicher (100 %), eine differentialdiagnostische Überlegung hatte er lediglich bis einschließlich zum 3. Informationsabschnitt angestellt.

Abbildung Nr. 26 zeigt ein Beispiel für ein sog. "Einzeldiagramm" eines Beurteilers von Film 1, das endgültig der Gruppe 2 zugeordnet wurde. Die einzelnen Gruppenbewertungen waren hier: 2-2-2, die "Flexibilität" ergab sich aus den 5 Kreuzungen bzw. Berührungen der einzelnen Diagnosen-Kurven (Mittelwert = 2,3), die "Sicherheit" ergab sich aus einer Gesamtdifferenz der subjektiven Sicherheitsgrade (SG%) zwischen der jeweiligen Haupt- und Zweitdiagnose von 330 (Mittelwert 641), die "Diagnosen-Summe" entstand aus den 27 einzelnen Diagnosen und Differentialdiagnosen, die während der gesamten diagnostischen Übung in Erwägung gezogen worden waren (Mittelwert = 21,3).

Abbildung Nr. 27 zeigt ein solches "Einzeldiagramm" eines Beurteilers von Film 2 mit der Zuordnung zur Gruppe 1. "Flexibilität" des Beurteilers bedeutet hier 0 Kreuzungen und Berührungen der einzelnen Diagnosen-Kurven (Gruppe 1, Mittelwert = 3,0). Die "Sicherheit" lautete 660 (Gruppe 1, Mittelwert = 491). Die "Diagnosen-Summe" setzte sich aus 14 Diagnosen zusammen (Gruppe 1, Mittelwert 18,4 Diagnosen).

Dieser Beurteiler zeigte kaum Veränderungen und Schwankungen hinsichtlich seiner diagnostischen Überlegungen, er war sich seiner Hauptdiagnose stets sehr sicher und stellte verhältnismäßig wenige, nur mit einem niedrigen subjektiven Sicherheitsgrad versehene differentialdiagnostische Überlegungen an.

Abbildung Nr. 28 stellt schließlich ein "Einzeldiagramm" eines Beurteilers von Film 1 mit der endgültigen Zuordnung zur Gruppe 2 dar. Aus der "Flexibilität" des Beurteilers (9 Kreuzungen und Berührungen der Diagnosen-Kurven), aus der "Sicherheit" (205 Gesamtdifferenz der subjektiven Sicherheitsgrade SG% zwischen Haupt- und Zweitdiagnosen) und aus der "Diagnosen-Summe" (34 während der diagnostischen Übung angeführte Diagnosen) ergaben sich die Gruppenbewertungen 2-2-2, woraus sich die endgültige Gruppe 2 ergab.



Zeichenerklärung siehe Anhang S.1 !

Abb. Nr. 25: Film 1

"Einzeldiagramm" eines Beurteilers,
der endgültig der Gruppe 1 zugeordnet
wurde.

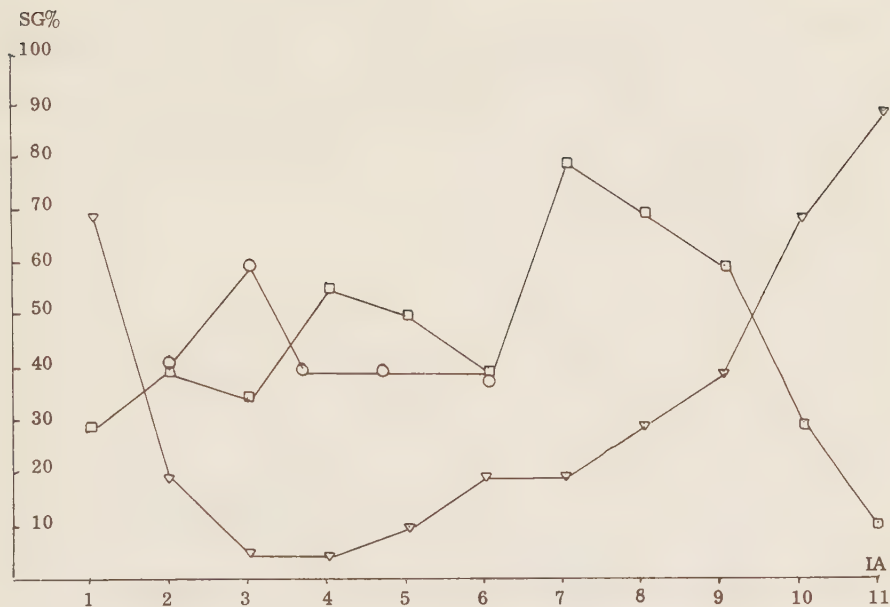


Abb. Nr. 26: Film 1

"Einzeldiagramm" eines Beurteilers, der
endgültig der Gruppe 2 zugeordnet wurde.

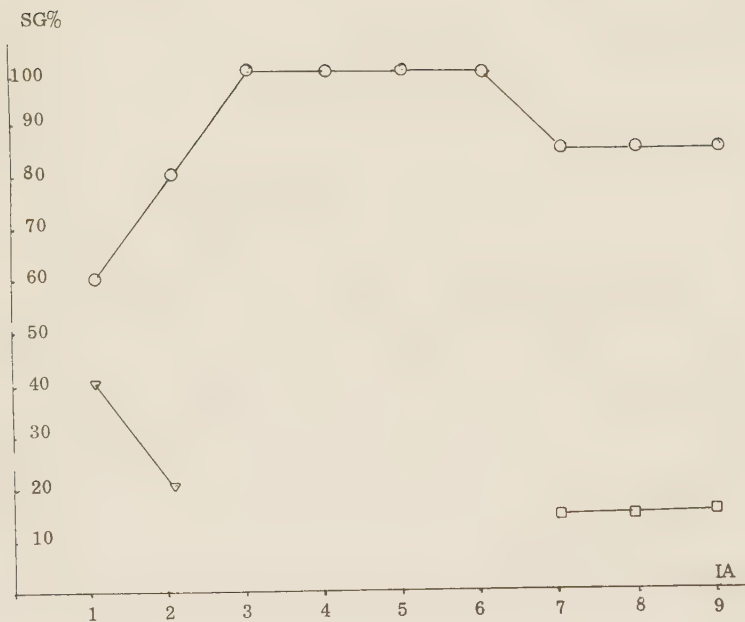


Abb. Nr. 27: Film 2

"Einzeldiagramm" eines Beurteilers, der
endgültig der Gruppe 1 zugeordnet wurde.

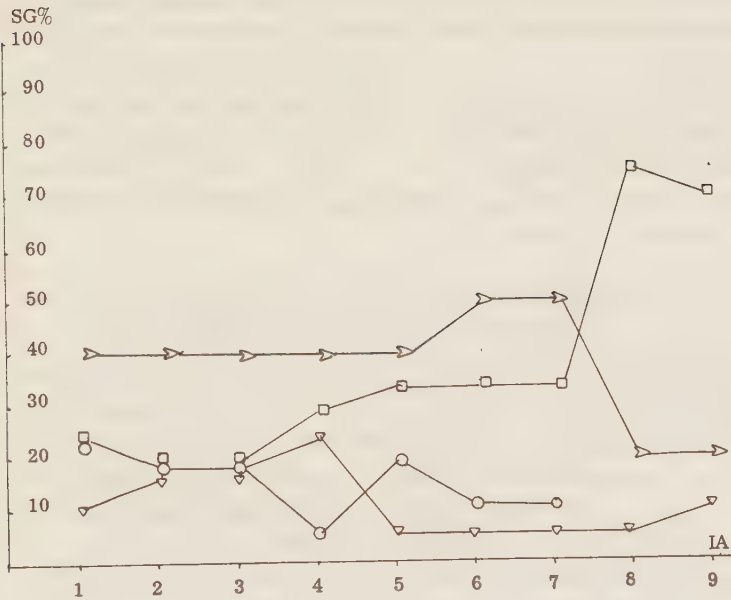


Abb. Nr. 28: Film 2

"Einzeldiagramm" eines Beurteilers, der
endgültig der Gruppe 2 zugeordnet wurde.

Der Beurteiler änderte seine diagnostischen Überlegungen während der Fernsehaufzeichnung verhältnismäßig rasch, er war sich seiner jeweiligen Hauptdiagnosen nicht sehr sicher und er stellte viele differentialdiagnostische Überlegungen an.

b) Die Beziehung zwischen der endgültigen Diagnose und dem diagnostischen Verhalten eines Beurteilers

Es sollte nun überprüft werden, ob gewisse Zusammenhänge zwischen der endgültigen Diagnose und dem diagnostischen Verhalten, charakterisiert durch die oben geschilderten Merkmale ("Flexibilität", "Sicherheit" und "Diagnosen-Summe") eines Beurteilers bestehen.

Nach dem Vierfelder-Test ließen sich mit statistischer Signifikanz ($p < 0,01$) folgende Zusammenhänge feststellen:

Beurteiler, die eine affektive Erkrankung als endgültige Diagnose gestellt hatten ("A", Film 2) zeigten in ihrem diagnostischen Verhalten eine "hohe Flexibilität", eine "niedrige Sicherheit" und eine "hohe Diagnosen-Summe" (= Beurteilergruppe 2). Beurteiler hingegen, die eine schizophrene Erkrankung als endgültige Diagnose gestellt hatten ("S", Film 2), zeigten in ihrem diagnostischen Verhalten eine entsprechend "niedrige Flexibilität", eine "hohe Sicherheit" und eine "niedrige Diagnosen-Summe" (= Beurteilergruppe 1). Beurteiler also, die ihre diagnostischen Überlegungen während der Patientenexploration rasch und häufig änderten ("hohe Flexibilität"), die ferner für ihre jeweiligen Hauptdiagnosen einen nur kleinen subjektiven Sicherheitsgrad SG% aufbrachten ("niedrige Sicherheit") und die relativ viele differentialdiagnostische Überlegungen ("hohe Diagnosen-Summe") angestellt hatten, neigten mit statistischer Signifikanz dazu, eine affektive Erkrankung zu diagnostizieren.

Beurteiler, die in ihren diagnostischen Überlegungen nur wenige Schwankungen zeigten ("niedrige Flexibilität"), die ferner für ihre diagnostischen Hauptentscheidungen während der Fernsehaufzeichnungen jeweils einen hohen subjektiven Sicherheitsgrad angegeben hatten ("hohe Sicherheit") und die relativ wenige differentialdiagnostische Überlegungen ("niedrige Diagnosen-Summe") angestellt hatten, neigten schließlich mit statistischer Signifikanz dazu ($p < 0,01$), eine schizophrene Erkrankung zu diagnostizieren.

Weiterhin ließen sich mit Hilfe des Vierfelder-Tests mit statistischer

Signifikanz folgende Feststellungen machen ($p < 0,01$):

Psychiater, die eine neurotische Erkrankung ("N", Film 1) endgültig diagnostiziert hatten, zeigten in ihrem diagnostischen Verhalten eine "hohe Flexibilität", d.h. sie änderten ihre diagnostischen Überlegungen verhältnismäßig rasch, ferner wiesen sie eine "niedrige Sicherheit" auf, d.h. sie hatten für ihre jeweiligen Hauptdiagnosen jeweils nur einen kleinen subjektiven Sicherheitsgrad SG% angegeben, außerdem stellten sie relativ viele differentialdiagnostische Überlegungen an und wiesen somit eine "hohe Diagnosen-Summe" (= Beurteilergruppe 2) auf.

Psychiater hingegen mit einer "niedrigen Flexibilität", mit einer "hohen Sicherheit" und einer "niedrigen Diagnosen-Summe" in ihrem diagnostischen Verhalten (= Beurteilergruppe 1) diagnostizierten mit statistischer Signifikanz häufiger eine Psychose ("P", Film 1).

c) Die Beziehung zwischen den Persönlichkeitseigenschaften der Psychiater und deren diagnostischem Verhalten

Die einzelnen Persönlichkeitsmerkmale der Beurteiler nach dem Freiburger Persönlichkeitsinventar FPI-A ergaben schließlich in einer weiteren Überprüfung unter Anwendung des χ^2 -Tests (Vierfelder-Test) keinerlei statistisch signifikante Korrelation mit dem diagnostischen Verhalten im oben definierten Sinne ("Flexibilität", "Sicherheit", "Diagnosen-Summe").

Speziell auch bei jenen Persönlichkeitsmerkmalen, die gewisse statistisch signifikante Unterschiede hinsichtlich der endgültigen Diagnosen aufwiesen (vgl. S. 58), konnten keine Zusammenhänge mit dem diagnostischen Verhalten eines Psychiaters gefunden werden.

d) Die Beziehung zwischen der klinisch-psychiatrischen Erfahrung der Beurteiler und deren diagnostischem Verhalten

Entsprechend wie unter Punkt c) konnten mit Hilfe des χ^2 -Tests (Vierfelder-Test) keinerlei Zusammenhänge zwischen der klinisch-psychiatrischen Erfahrung der Beurteiler und deren diagnostischem Verhalten festgestellt werden. Es galt folgende Ausnahme: Beurteiler von Film 1, die über mehr als 2 1/2 Jahre klinischer Erfahrung in der Psychiatrie verfügten, wiesen in ihrem diagnostischen Verhalten eine relativ "hohe Diagnosen-Summe" auf, d.h. sie

stellten mit statistischer Signifikanz ($p < 0,01$) mehr differentialdiagnostische Überlegungen an als Psychiater mit weniger als 2 1/2 Jahren klinischer Erfahrung in der Psychiatrie.

e) Vergleich des diagnostischen Verhaltens von Beurteilern beim ersten und beim zweiten diagnostischen Interview

Es waren insgesamt nur 18 Ärzte, die sowohl an der ersten als auch an der zweiten diagnostischen Übung teilgenommen hatten. Aus einer Berechnung ergab sich, daß das diagnostische Verhalten, charakterisiert durch die oben beschriebenen Merkmale "Flexibilität", "Sicherheit" und "Diagnosen-Summe", bei 55,5 % aller Psychiater für beide Fernsehaufzeichnungen gleich blieb, während sich bei 44,5 % aller Beurteiler beider Interviews dieses diagnostische Verhalten änderte.

DISKUSSION

Diese Untersuchung wurde durchgeführt, weil überraschend wenig über das Beurteilungsverhalten und die diagnostischen Gewohnheiten von Psychiatern bekannt ist. Sie kann insofern als Reliabilitätsstudie bezeichnet werden, als eine verhältnismäßig große Anzahl von psychiatrischen Beurteilern zwei Patienten während einer klar umschriebenen Krankheitsphase diagnostizierten. Die einzige Informationsquelle waren die Fernsehaufzeichnungen, die allen Diagnostikern gleiche Daten zur Verfügung stellten.

Die vierstelligen ICD-Diagnosen faßten wir zu den einzelnen Hauptgruppen zusammen, da durch das Auftreten von sehr vielen Untergruppen zu große diagnostische Uneinstimmigkeiten und zu viele Änderungen der Hauptdiagnosen im Laufe der diagnostischen Interviews zu erwarten gewesen wären, die den tatsächlichen diagnostischen Schwankungen der betreffenden Beurteiler nicht immer entsprochen hätten. Eine große und in unserem Sinne damit wirkliche diagnostische Schwankung stellt beispielsweise die Überlegung dar, ob es sich um eine Schizophrenie oder um eine schizoide Neurose handelt, nicht jedoch die Überlegung, ob es sich um eine depressive oder um eine hypochondrische Neurose handelt. Auf die ganz feinen diagnostischen Differenzierungen sind wir in unserer Untersuchung nicht weiter eingegangen.

Trotz der Ergebnisse von GAURON und DICKINSON (1969), von SANDIFER (1970) und von KENDELL (1973), nämlich daß diagnostische Entscheidungen in der Psychiatrie verhältnismäßig früh getroffen werden, dauerten unsere diagnostischen Interviews relativ lange (44 bzw. 48 Minuten), nicht zuletzt auch, um die frühen diagnostischen Entscheidungen der Beurteiler mit deren endgültigen Diagnosen am Schluß einer Exploration vergleichen zu können. Die Ergebnisse der oben genannten Autoren wurden hierbei weitgehend bestätigt. In unserer Untersuchung wurden die diagnostischen Entscheidungen im allgemeinen bereits nach der ersten Filmunterbrechung (4 bzw. 4 1/2 Minuten) gefällt.

Gleichzeitig konnten wir feststellen, daß auch die subjektive Sicherheit der Beurteiler, die sie für ihre Diagnosen angaben, global gesehen nahezu unverändert blieb. Es war also am Ende der diagnostischen Interviews die subjektive Sicherheit der Psychiater hinsichtlich ihrer endgültigen Diagnose nicht größer als nach der ersten Filmunterbrechung. Freilich darf nicht vergessen

werden, daß diese Feststellung global gesehen für alle Beurteiler zusammen gemacht wurde. In einzelnen Fällen boten sich oftmals recht unterschiedliche Bilder bezüglich dieser subjektiven Sicherheiten.

THORNDIKE (1920) prägte den Begriff "halo effect", um damit auf einen permanenten Fehler, den er bei der Beurteilung von Persönlichkeitsstörungen herausfand, hinzuweisen: "Ein Beurteiler neigt dazu, von Anfang an eine generelle Entscheidung zu treffen, dahingehend daß eine Person beispielsweise allgemein gut oder schlecht sei, und schreitet demgemäß immer mehr in der Differenzierung dieser Merkmale fort, ohne noch ausreichendes Beweismaterial dafür zu haben." (THORNDIKE, 1920).

In diesem Sinn ist wohl auch die Tatsache zu verstehen, daß bei einem sehr hohen Prozentsatz der Beurteiler die anfängliche diagnostische Entscheidung identisch ist mit deren endgültiger Diagnose, daß alle weiteren Informationen während der diagnostischen Interviews für sie also keine wesentlichen Änderungen in der Diagnose mehr zu bringen scheinen. Wenn man also einmal eine Entscheidung trifft, besteht offenbar eine gewisse Erwartungshaltung, diesen ersten Eindruck bestätigt zu sehen. Es können bestimmte Merkmale von einem Beurteiler in eine andere Person hineinprojiziert werden, die dann mehr oder weniger unbewußt zu einer Persönlichkeitseinstufung führen, ohne daß dies für irgendeinen anderen Beurteiler klar ersichtlich oder beweiskräftig wäre.

Darüberhinaus überraschte die Feststellung, daß die Symptomwahrnehmung offenbar keinen Einfluß auf die Diagnose auszuüben scheint. Es sei nochmals auf die Ergebnisse von KREITMAN, SAINSBURY, MORRISSEY, TOWERS und SCRIVENER (1961) hingewiesen, die in ihrer Untersuchung herausfanden, daß insgesamt mehr Übereinstimmung unter den Diagnostikern, die den gleichen Patienten im Abstand von einigen Tagen sahen, hinsichtlich der Diagnose erreicht wurde, als bei den Bewertungen individueller Symptome. Die Variabilität auf Symptomebene hat also keinen statistisch relevanten Einfluß auf die Wahl der Diagnose.

Die einzige diesbezügliche Ausnahme stellten wir bei Film 1 unter der Merkmalsgruppe "Verlaufsbesonderheiten" fest. Die Psychiater P, die eine Psychose diagnostizierten, hatten auf ihren AMP-Belegblättern wesentlich mehr psychopathologische Symptome (hier speziell "abends besser") aus der genannten Merkmalsgruppe angegeben, als die Psychiater N, die eine Neurose diagnostizierten. Da jedoch am häufigsten eine endogene Depression als Psychose diagnostiziert

worden war und bei diesem Krankheitsbild abends ein allgemeines Anheben der Stimmungslage des Patienten zu erwarten ist, dürfte diese Feststellung kein ausdrücklicher Hinweis auf ein sich gegenseitiges Beeinflussen von Symptombemerkung und endgültiger Diagnose sein. Man könnte hier im Gegenteil vielleicht spekulativ annehmen, daß die eventuell vorgefaßte Diagnose einer endogenen Depression einen gewissen Einfluß auf die Art der Symptombemerkung hatte.

Erinnert man nun an die Aussage von SHEPHERD, BROOKE, COOPER und LIN (1968), daß eine Diagnose als Zusammenfassung vieler einzelner Beobachtungen zu verstehen und somit auch verlässlicher als ihre Bestandteile sei, erscheint unser gefundenes Phänomen verstehbar. Auch ein gewisser "geistiger Auswahlvorgang" von Seiten der Beurteiler, nach dem bestimmte Informationen unbewußt ganz einfach übergangen werden können, da vielleicht nicht alle Stimuli von allen Beurteilern gleichzeitig wahrgenommen werden, haben nach Meinung der oben genannten Autoren vielleicht unterschiedliche Symptombemerkungen zur Folge, ohne dabei am Ende unterschiedliche diagnostische Entscheidungen zu verursachen.

Die unterschiedliche Symptombemerkung der Beurteiler dürfte auch darauf beruhen, daß das Verhalten der Patienten vielfach anders ist als die Informationen, die sie in Form von Fragenbeantwortung von sich geben. Die vom Patienten subjektiv empfundenen Symptome und die vom Beurteiler festgestellte Psychopathologie müssen erfahrungsgemäß nicht miteinander identisch sein. Denkt man beispielsweise an den Patienten des zweiten diagnostischen Interviews, so ergab sich für die Beurteiler das Problem, zwischen einer vom Patienten subjektiv überaus stark empfundenen Religiosität (hier der Glaube an Guru Maharadshi) und vielleicht einem religiösen Wahn zu differenzieren. Je nachdem ob nun ein Beurteiler das eine oder das andere Merkmal für zutreffend hielt, ergaben sich in der Summe unterschiedliche Ergebnisse bezüglich der Symptombemerkung.

Der von THORNDIKE (1920) geprägte Begriff "halo effect" gilt sowohl für die Diagnose im Ganzen als auch für die detaillierte Symptombemerkung.

SHEPHERD, BROOKE, COOPER und LIN (1968) meinten dazu: "Wenn man von Anfang an einer Exploration vermutet, daß eine Schizophrenie die endgültige Diagnose ist, so wird die Beobachtung, daß der Patient ziemlich langsam und ruhig ist und wenige Emotionen zeigt, bereits als ein Verflachen der Affekte bewertet und somit die anfängliche Diagnose bestätigt. Von einem anderen Beobachter des gleichen Interviews, der aus anderen Gründen früher die Diagnose einer Depression stellte, kann das gleiche Verhalten als "depressive Apathie"

ausgelegt werden. Somit kann das gleiche Verhalten eines Patienten für verschiedene Diagnosen angewandt werden."

Die diagnostischen Erwartungen eines Beurteilers können also einen wichtigen Einfluß auf die Folgerungen, die er aus seinen Beobachtungen zieht, nehmen.

Die psychiatrische Diagnose ist schließlich auch deshalb so wenig verläßlich, weil viele Symptome verschiedenen Krankheiten so ähnlich sind und eben je nach den diagnostischen Erwartungen eines Beurteilers verschieden interpretiert werden können.

Vielleicht würde das Zusammenfügen von Symptomen zu Syndromen diese Fehlerquelle psychiatrischer Diagnostik ausschalten helfen und somit auch die Reliabilität verbessern.

Aufgrund früherer Untersuchungen nahmen wir stark an, daß die verschiedenen Persönlichkeitsmerkmale der einzelnen Beurteiler für die große Variabilität in der Diagnostik eine wichtige Rolle spielen.

Die mit Hilfe des Freiburger Persönlichkeitsinventares gefundenen Zusammenhänge zwischen einzelnen Persönlichkeitsmerkmalen und der endgültigen Diagnose bestimmter Beurteiler sind zunächst sehr interessant: Beurteiler mit einer relativ hohen Aggressivität diagnostizierten beispielsweise eher eine Psychose als Beurteiler mit einer relativ geringen Aggressivität, die wiederum eher eine Neurose diagnostizierten usw. . Es können hier viele psychologische Überlegungen angestellt werden, um mögliche Interpretationen für diese Ergebnisse zu finden. Eine endgültige Erklärung dieser diffizilen Zusammenhänge läßt sich allerdings anhand der FPI-Kurzform und der sehr begrenzten Fallzahl nicht geben. Es darf auch nicht übersehen werden, daß die Beurteiler alle psychiatrisch geschult waren und somit wohl auch nicht ganz frei und objektiv bei der Beantwortung der FPI-Standardfragen vorgehen konnten.

Wir wandten in unserer Untersuchung das AMP-System an, um zu vermeiden, daß verschiedene Diagnostiker die gleichen Worte gebrauchen, um verschiedene Phänomene zu beschreiben oder umgekehrt gleiche psychopathologische Phänomene mit unterschiedlichen Ausdrücken beschrieben werden. Es sollte auf diese Weise eine gewisse Übereinstimmung zumindest in der Terminologie erzielt werden. Nachdem alle Teilnehmer mit dem AMP-System vertraut waren, konnten diese geschilderten Fehlerquellen weitgehend ausgeschaltet und die Untersuchungsergebnisse nachprüfbarer gemacht werden.

Mittels des AMP-Belegblattes "Psychischer Befund" wurden jedoch der somati-

sche Befund, die Ätiologie der Erkrankung und die spezielle Anamnese der Patienten nicht erfaßt. Hierbei wurde in erster Linie auf die deskriptive Komponente des psychopathologischen Befundes eingegangen. In der Diagnostik dürften, wenn auch nicht nachweisbar, auch die anderen oben geschilderten Komponenten, die während der Exploration schließlich auch angesprochen worden waren, eine gewisse Rolle gespielt haben. Eine mehrdimensionale Diagnostik (hier mit Hilfe aller dokumentierten Befundbögen des AMP-Systems) hätte eine genauere Überprüfung der Beurteilerübereinstimmung in der Diagnostik und der Symptomebeobachtung ermöglicht. Vielleicht hätte sich dadurch auch ein deutlicherer Zusammenhang zwischen der endgültigen Diagnose der Beurteiler und deren Symptomwahrnehmung ergeben.

Die Feststellung, daß erfahrener Psychiater offenbar mehr Symptome aufdecken als ihre jüngeren Kollegen, stimmt mit den Untersuchungsergebnissen von COPELAND, COOPER, KENDELL und GOURLAY (1971) überein. Dies könnte vielleicht so verstanden werden, daß im Laufe einer psychiatrischen Facharztausbildung ein gewisser diagnostischer Scharfsinn erst geprägt oder ausgebildet werden muß, um den Arzt für die vielfältige psychiatrische Symptomatik gewissermaßen zu "sensibilisieren", bis er einen detaillierten psychopathologischen Befund aufstellen kann.

Die Wahrnehmungswelt eines Diagnostikers ist also sehr subjektiv und spiegelt zum Teil seine eigene Situation wieder. Es zeigt sich auch, daß die Haltung eines Beurteilers gegenüber dem, was als "krank" und was als "gesund" zu bezeichnen ist, irgendwie mit dem Erfahrungsalter und der Länge der Ausbildung eines Psychiaters verknüpft zu sein scheint, ohne daß sich hierfür eine offensichtliche Erklärung finden läßt.

Es ist allerdings schon von großem Interesse, allein das Vorhandensein eines solchen Trends zu sehen, nicht zuletzt auch, um zu weiteren Nachforschungen anzuregen.

Interessant war in unserer Untersuchung auch jene Feststellung, daß die weniger erfahrenen Psychiater dazu neigten, eine affektive bzw. eine neurotische Erkrankung zu diagnostizieren als ihre erfahreneren Kollegen. Letztere neigten eher dazu, eine schizophrene Erkrankung bzw. eine Psychose ganz allgemein zu diagnostizieren.

Da nur zwei Patienten beurteilt worden waren, läßt sich über die "Art" dieser Unterschiedlichkeiten keine weitere Aussage machen und inwieweit vielleicht

gewisse eigene Emotionen von Seiten der Ärzte mitspielen, kann nicht beurteilt werden.

Das Problem eventueller unterschiedlicher Lehrmeinungen, wie es in der Arbeit von COPELAND, COOPER, KENDELL und GOURLAY (1971) zu erkennen war, konnte in dieser Untersuchung weitgehend vermieden werden, da die Psychiater fast alle an der Universitäts-Nervenklinik München ausgebildet worden waren bzw. sich noch in ihrer Facharztausbildung befanden.

Eine Schwäche in der Methodik bestand darin, daß die Beurteiler zwar Differentialdiagnosen, jedoch keine Zusatzdiagnosen stellen durften. Irgendwelche Diagnosenkombinationen waren somit nicht möglich, was den subjektiven Sicherheitsgrad der Beurteiler bezüglich ihrer Hauptdiagnosen erheblich beeinflußt haben dürfte. Die Autoren BOCHNIK, HELMCHEN, HIPPIUS, KNÜPPEL, KULENKAMPFF, LAUTER, MEYER, MÜLLER, WIESER und WINKLER (1970) stellten nämlich in einer großen Untersuchung an sechs deutschen psychiatrischen Kliniken fest, daß sich die Brauchbarkeit der psychiatrischen Diagnosen der WHO-Diagnosen-Klassifikation erst deshalb als so befriedigend erwiesen, als gewisse Diagnosen-Kombinationen erlaubt waren.

Schließlich soll noch einmal darauf hingewiesen werden, daß die Patienten verhältnismäßig schwierig zu diagnostizieren waren. Die diagnostische Variabilität war bereits von Anfang an sehr groß. Bei eindeutiger zu klassifizierenden Patienten hätte möglicherweise eine höhere Übereinstimmung unter den Psychiatern erzielt werden können. Wir betrachten die Ergebnisse jedoch als sehr wichtig, da sie repräsentativ sind für die alltäglichen diagnostischen Probleme einer psychiatrischen Klinik.

Das Hauptziel dieser Arbeit war es schließlich, die diagnostischen Überlegungen eines Psychiaters, seinen diagnostischen Prozess, näher zu beschreiben. Es zeigte sich dabei, daß die diagnostischen Überlegungen von Psychiatern statistisch nur sehr schwer einzuordnen sind.

Anhand der "Einzeldiagramme" entstand der Eindruck zweier verschiedener Gruppen. Wir entwickelten drei Indices, wonach sich diese Gruppen auch zahlenmäßig erfassen ließen. Beurteiler mit relativ vielen diagnostischen Schwankungen, wir nannten es "Flexibilität", mit einer vergleichsweise niedrigen subjektiven Sicherheit ("Sicherheit") und vielen differentialdiagnostischen Überlegungen ("Diagnosen-Summe") boten ein anderes graphisches Bild ihres diagnostischen Verhaltens als Beurteiler mit den entgegengesetzten Merkmalen.

Freilich gehen diese Merkmale kontinuierlich ineinander über, eine scharfe Abgrenzung dieser beiden gefundenen Gruppen voneinander ist nicht möglich. Dennoch lassen sich die beiden Beurteilergruppen in ihrem diagnostischen Vorgehen voneinander unterscheiden durch bestimmte Gesetzmäßigkeiten, eben ausgedrückt durch die von uns entwickelten Indices.

Mit Hilfe dieser graphischen Darstellungen versuchten wir, die diagnostischen Überlegungen von Psychiatern objektivierbar zu machen, ohne jedoch den Anspruch erheben zu wollen, über ein detaillierteres diagnostisches Verhalten von Psychiatern eine Aussage machen zu können.

Wir versuchten beispielsweise in den Abbildungen Nr. 25 und Nr. 26 zwei solcher "Einzeldiagramme" gegenüberzustellen.

Die Abbildung Nr. 25 stellt den diagnostischen Vorgang eines sehr erfahrenen Psychiaters dar, der während der ersten drei Informationsabschnitte differentialdiagnostisch noch eine endogene Depression in Erwägung gezogen hatte. Seine Hauptdiagnose, eine Neurose, erhielt ab dem vierten Informationsabschnitt einen subjektiven Sicherheitsgrad von 100 % und wurde bis zum Schluß der Untersuchung nicht mehr geändert.

Dieser Beurteiler zeigte nach den von uns entwickelten Indices eine kleine "Flexibilität", eine hohe "Sicherheit" und eine kleine "Diagnosen-Summe". Mit der kleinen "Diagnosen-Summe" gingen speziell bei diesem Beurteiler sehr viele Symptombeobachtungen einher, was ein Hinweis auf eine gewisse diagnostische Genauigkeit und auch Sicherheit sein dürfte.

Die endgültige Diagnose des betreffenden Beurteilers, eine Neurose, stimmt allerdings mit der von uns festgestellten Tendenz nicht überein, daß speziell bei Film 1 eine große klinische Erfahrung im allgemeinen mit der endgültigen Diagnose einer schizophrenen Erkrankung einherging.

Die Abbildung Nr. 26 stellt den diagnostischen Prozess eines Beurteilers mit weniger als 2 1/2 Jahren klinisch-psychiatrischer Erfahrung dar. Seine Hauptdiagnose nach dem ersten Informationsabschnitt war eine neurotische Erkrankung, dann war er bis zum 9. Informationsabschnitt mit relativ großer Sicherheit davon überzeugt, daß eine endogene Depression bei der zu beurteilenden Patientin vorliegt, bis er am Schluß der Untersuchung mit einem subjektiven Sicherheitsgrad von 90 % auf seine anfangs erteilte Hauptdiagnose wieder zurückkam. Zwischendurch stellte er differentialdiagnostische Überlegungen in Richtung einer schizophrenen Erkrankung an.

Dieser Beurteiler zeigte gemäß unseren Indices eine große "Flexibilität", eine verhältnismäßig kleine "Sicherheit" und eine hohe "Diagnosen-Summe". Der

Beurteiler hatte auf seinen AMP-Belegblättern nur wenige Symptome vermerkt, was wiederum mit seinen häufigen differentialdiagnostischen Überlegungen hinsichtlich einer gewissen Unsicherheit korrelieren könnte.

Die endgültige Diagnose einer neurotischen Erkrankung stimmt mit der von uns festgestellten Tendenz, daß weniger erfahrene Psychiater in Film 1 eher eine neurotische Erkrankung diagnostizierten, überein.

Die im Anhang abgebildeten "Einzeldiagramme" sämtlicher Beurteiler sind geordnet nach ihrem diagnostischen Vorgehen.

Gewisse Gesetzmäßigkeiten, nach denen sich die Beurteiler beider gefundenen Gruppen unterscheiden, dürften wohl klar ersichtlich sein, ebenso wie die allmähliche Kontinuität, in der diese beiden Beurteilergruppen ineinander übergehen.

Es fiel schließlich auf, daß Psychiater mit ähnlichen diagnostischen Gepflogenheiten (eingeteilt in Gruppe 1 und in Gruppe 2) auch ähnliche Diagnosen bevorzugt stellten. Schizophrenie bzw. Psychose wurden speziell von den weniger "flexiblen" und den "sicheren" Beurteilern diagnostiziert, die auch relativ wenige differentialdiagnostische Überlegungen angestellt hatten. Neurotische und affektive Erkrankungen diagnostizierten bevorzugt die "flexibleren" und weniger "sicheren" Beurteiler, die auch relativ viele Differentialdiagnosen in Erwägung gezogen hatten.

Bei dem Vergleich der diagnostischen Gewohnheiten von Beurteilern beider Interviews stellten wir allerdings fest, daß die Psychiater ihr diagnostisches Verhalten von Fall zu Fall änderten. Wenn dies aufgrund der relativ geringen Beurteiler- und Patientenzahl vielleicht auch nur Zufallsbefunde waren, so läßt sich doch eine gewisse Parallele zu den Untersuchungsergebnissen von GAURON und DICKINSON (1966) feststellen. Die Autoren fanden heraus, daß die Ärzte ihr diagnostisches Verhalten, wiederum klassifiziert durch bestimmte von den Autoren aufgestellte Indices, während einer diagnostischen Exploration und auch bei mehreren unterschiedlichen Interviews mehrfach änderten. Ein pauschales Zuordnen bestimmter diagnostischer Verhaltensweisen scheint also ganz allgemein ein sehr schwieriges Unterfangen zu sein. Ein und derselbe Psychiater kann seine diagnostischen Verhaltensweisen ständig ändern, ohne irgendwelche Gesetzmäßigkeiten dabei aufzuweisen.

Weitaus komplizierter noch dürfte es sein, gewisse Beziehungen zwischen den Persönlichkeitseigenschaften von Psychiatern und deren diagnostischem Verhalten herzustellen.

Unsere negativen Ergebnisse bei der Überprüfung solcher Zusammenhänge überraschen nicht weiter. Eventuelle tiefenpsychologische Aspekte mußten ganz außer Acht gelassen werden und die Feststellung von Persönlichkeitsmerkmalen nach der FPI-Kurzform war sicher nicht ausführlich genug für solch eine feine und komplizierte Untersuchung.

Die Forschung auf diesem Gebiet sei einer anderen wissenschaftlichen Untersuchung mit einer diesbezüglich verfeinerten Methodik überlassen.

Schließlich sei noch einmal darauf hingewiesen, daß diese allgemeine experimentelle Situation Gefahr läuft, einen höchst komplizierten Vorgang, den diagnostischen Prozess eines psychiatrischen Beurteilers, zu vereinfachen.

Dennoch brachte diese Untersuchung sehr interessante Ergebnisse, die zu weiteren Nachforschungen anregen sollten.

ZUSAMMENFASSUNG

In einer ausführlichen literarischen Übersicht über die bisherigen Reliabilitätsstudien wurde die Problematik der Verlässlichkeit psychiatrischer Diagnostik aufgezeigt.

Zwei Fernsehaufzeichnungen - psychiatrische Explorationen zweier schwierig zu diagnostizierender Patienten - von 48 bzw. 44 Minuten Dauer und 11 bzw. 9 Filmunterbrechungen wurden insgesamt von 66 Psychiatern (Film 1 von 31 Psychiatern, Film 2 von 35 Psychiatern) unter Anwendung der WHO-Diagnosen-Klassifikation (ICD) und des AMP-Systems (3. Psych. 3) beurteilt. Die Psychiater stellten nach jeder Filmunterbrechung Diagnosen und Differentialdiagnosen und gaben jeweils einen subjektiven Sicherheitsgrad in Prozent für diese Diagnosen an. Die beobachtete Symptomatik wurde nach jedem Informationsabschnitt auf den AMP-Belegblättern festgehalten. Am Schluß der Interviews füllten die Beurteiler selbst die Halbform FPI-A des Freiburger Persönlichkeitsinventares aus.

Aus der Untersuchung ließ sich entnehmen, daß diagnostische Entscheidungen meist sehr früh getroffen werden (in unserem Fall bereits nach 4 bzw. 4 1/2 Minuten), daß die anfängliche Diagnose im allgemeinen mit der endgültigen Diagnose am Schluß der Exploration übereinstimmt und der subjektive Sicherheitsgrad global gesehen für alle Beurteiler und für alle diagnostischen Hauptgruppen, die wir aus den vierstelligen ICD-Diagnosen gebildet hatten, von Anfang an gleich verläuft.

Die Interpretation der Testergebnisse des Freiburger Persönlichkeitsinventares ergab eine als "gesund" und "normal" zu bezeichnende Population der psychiatrischen Teilnehmer unserer Untersuchung.

Einzelne Persönlichkeitsmerkmale der Beurteiler nach dem Freiburger Persönlichkeitsinventar zeigten eine statistisch signifikante Korrelation mit den endgültig gestellten Diagnosen dieser Beurteiler: "hohe Aggressivität" korrelierte in Film 1 mit der endgültigen Diagnose einer neurotischen Erkrankung. Beurteiler mit einer "starken Extraversion" neigten in Film 1 dazu (keine statistische Signifikanz), eine Psychose zu diagnostizieren, Beurteiler mit einer "starken Introvertiertheit" neigten entsprechend mehr dazu, eine Neurose zu diagnostizieren. Beurteiler von Film 2 mit einer "hohen Erregbarkeit" tendierten dazu (keine statistische Signifikanz), eine schizophrene Erkrankung zu

diagnostizieren, Beurteiler mit einer "niedrigen Erregbarkeit" diagnostizierten eher eine affektive Erkrankung.

Weiterhin konnte die Feststellung gemacht werden, daß die Variabilität auf Symptomebene keinen statistisch relevanten Einfluß auf die Wahl der Diagnose hat. Die Symptombeobachtungen der Beurteiler, festgehalten auf den AMP-Befundbögen, zeigten keinen Zusammenhang mit den endgültigen Diagnosen der Beurteiler. Lediglich eine Ausnahme ergab sich unter der Merkmalsgruppe "Verlaufsbesonderheiten" bei Film 1: Beurteiler, die endgültig eine Psychose diagnostiziert hatten, hatten wesentlich mehr Symptome aus der genannten Merkmalsgruppe (speziell das Symptom "abends besser") beobachtet als die Beurteiler, die eine neurotische Erkrankung diagnostiziert hatten.

Beurteiler mit mehr als 2 1/2 Jahren klinisch-psychiatrischer Erfahrung beobachteten mit statistischer Signifikanz häufiger Symptome als Beurteiler mit weniger als 2 1/2 Jahren Erfahrung in der Psychiatrie. Bestimmte Merkmalsgruppen waren bei beiden Filmen verschieden stark betroffen.

Zur Beschreibung des diagnostischen Prozesses der Psychiater wurden für jeden einzelnen Beurteiler sog. "Einzeldiagramme" angefertigt, in denen die verschiedenen Diagnosen (Haupt- und Differentialdiagnosen über sämtliche Informationsabschnitte hinweg) in Abhängigkeit von den jeweils dazugehörigen subjektiven Sicherheitsgraden graphisch dargestellt sind.

Es entstand hierbei der Eindruck zweier verschiedener Beurteilergruppen. Wir entwickelten drei Indices, wonach sich diese beiden Gruppen auch zahlenmäßig erfassen ließen:

Die "Flexibilität" (Zahl der Überschneidungen und Berührungen der Diagnosenverläufe auf den "Einzeldiagrammen"), die "Sicherheit" (Differenz der subjektiven Sicherheitsgrade zwischen den jeweiligen Haupt- und Zweitdiagnosen sämtlicher Informationsabschnitte) und die "Diagnosen-Summe" (Summe aller Diagnosen und Differentialdiagnosen, die während der ganzen diagnostischen Übung genannt wurden).

Es ließ sich mit Hilfe dieser graphischen Darstellungen ein statistisch signifikanter Zusammenhang zwischen dem diagnostischen Verhalten der Beurteiler, ausgedrückt durch die oben geschilderten Indices, und deren endgültiger Diagnose feststellen.

Psychiater der Beurteilergruppe 1 ("niedrige Flexibilität", "hohe Sicherheit", "kleine Diagnosen-Summe") diagnostizierten vorwiegend eine Psychose (Film 1) bzw. eine Schizophrenie (Film 2), Psychiater der Beurteilergruppe 2 ("hohe

Flexibilität", "niedrige Sicherheit", "hohe Diagnosen-Summe") diagnostizierten vorwiegend eine Neurose (Film 1) bzw. eine affektive Erkrankung (Film 2).

Bei dem Vergleich des diagnostischen Verhaltens von Psychiatern, die an beiden diagnostischen Übungen teilgenommen hatten, war zu erkennen, daß die Beurteiler ihr diagnostisches Verhalten von einem zum anderen Patienteninterview änderten. Feste diagnostische Gepflogenheiten, ausgedrückt durch die oben beschriebenen Indices, ließen sich den einzelnen Psychiatern demnach nicht zuordnen.

Es sollte hierfür eine eigene Studie durchgeführt werden, deren Ziel es ist, die Veränderbarkeit bzw. die Konstanz des diagnostischen Verhaltens von Psychiatern, die mehrere Patienten beurteilen, zu untersuchen.

LITERATURVERZEICHNIS

1. AMP-System
Manual zur Dokumentation psychiatrischer Befunde.
Herausgegeben von der Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie.
Springer-Verlag 1972
2. APA-Nomenklatur (American Psychiatric Association)
Diagnostic and Statistical Manual of Mental Disorders.
Washington, D. C. 1952
3. ASH, P.
J. Abnorm. Soc. Psychol. 44:272
1949
4. BARBIGAN, H. M., GARDNER, E. MILES, H. C. und ROMANO, J.
Diagnostic Consistency and Change in a Follow-up Study of 1215 Patients.
Amer. J. Psychiat., 121, 895-901
1965
5. BECK, A. T.
Amer. J. Psychiat. 191, 210-216
1962
6. BECK, A. T., WARD, C. H., MENDELSON, M., MOCK, J. E. und ERBAUGH, J. K.
Reliability of Psychiatric Diagnosis: A Study of Consistency of Clinical Judgements and Ratings.
Amer. J. Psychiat. 119, 351-357
1962
7. BEHRENDTS, K., FLEGEL, H., HELMCHEN, H., HIPPIUS, H., HÖFFKEN, K. D., SCHACHT, L. und SCHULTE, P. W.
Quantifizierung psychiatrischer Symptome unter transkulturellen Aspekten.
Sozialpsychiatrie 6, 66-72
1971
8. BOCHNIK, H. J., HELMCHEN, H., HIPPIUS, H., KNÜPPEL, H., KULENKAMPFF, C., LAUTER, H., MEYER, J. E., MÜLLER, H. W., WIESER, ST. und WINKLER, W. TH.
Zur Brauchbarkeit der psychiatrischen Diagnosen der WHO-Diagnosen-Klassifikation.
Nervenarzt, 41. Jg. Heft 1, 42-44
1970
9. COOPER, J. E.
Diagnostic Change in a Longitudinal Study of Psychiatric Patients.
Brit. J. Psychiat., 113, 129-142
1967
10. COOPER, J. E.
The Use of a Procedure for Standardising Psychiatric Diagnosis. In Psychiatric Epidemiology, eds. E. H. Hare und J. K. Wing, Oxford, University Press
1970

11. COOPER, J. E., KENDELL, R. E., GURLAND, B. J., SARTORIUS, N. und FARKAS, T. Cross-national Study of Diagnosis of Mental Disorders: Some Results from the First Comparative Investigation. Amer. J. Psychiat., 125, 10, Supp. 21-9, 1969
12. COOPER, J. E., KENDELL, R. E., GURLAND, B. J., SHARPE, L., COPELAND, J. R. M. und SIMON, R. Psychiatric Diagnosis in New York and London. Oxford University Press: London 1972
13. COPELAND, J. R. M., COOPER, J. E., KENDELL, R. E. und GOURLAY, A. J. Differences in Usage of Diagnostic Labels Amongst Psychiatrists in the British Isles. Brit. J. Psychiat., 118, 629-640 1971
14. Diagnoseschlüssel und Glossar psychiatrischer Krankheiten, Deutsche Übersetzung der internationalen Klassifikation der WHO: ICD (International Classification of Diseases), 8. Revision und des internationalen Glossars. 1970
15. ENGELSMANN, F., VINAR, O., PICHOT, P., HIPPIUS, H., GIBERTI, F., ROSSI, L. und OVERALL, J. E. International Comparisons of Diagnostic Patterns. Transcultural Psychiatric Research, 7, 130-137 1970
16. FOULDS, G. A. J. Ment. Sci., 101:851 1955
17. Freiburger Persönlichkeitsinventar (FPI) FAHRENBERG, J. und SELG, H. Handanweisung für die Durchführung und Auswertung. Verlag für Psychologie, Dr. C. J. Hogrefe, Göttingen 1970
18. GAURON, E. F. und DICKINSON, J. K. Diagnostic Decision Making in Psychiatry. I. Information Usage. II. Diagnostic Styles. Arch. Gen. Psychiat. 14, 225-237 1966
19. GAURON, E. F. und DICKINSON, J. K. The Influence of Seeing the Patient First on Diagnostic Decision Making in Psychiatry. Amer. J. Psychiat. 126, 199-205 1969

20. GURLAND, B. J., FLEISS, J. L.,
COOPER, J. E., KENDELL, R. E.
und SIMON, R. Cross-national Study of Diagnosis of
the Mental Disorders; Some Comparisons
of Diagnostic Criteria from the First
Investigation.
Amer. J. Psychiat., 125, 10 Supp. 30-39
1969

21. GURLAND, B. J., FLEISS, J. L.,
COOPER, J. E., SHARPE, L.,
KENDELL, R. E. und
ROBERTS, P. Cross-national Study of Diagnosis of
the Mental Disorders; Hospital Dia-
gnosis and Hospital Patients in New
York and London.
Comprehensive Psychiat., 11, 18-25
1970

22. HELMCHEN, H., HIPPIUS, H.
und STÜRZBECHER, M. Ergebnisse, Kasuistik und Diskussion
zur psychiatrischen WHO-Diagnosen-
Statistik an vier Berliner Nervenklini-
ken.
Nervenarzt, 43, 427-432
1972

23. HUNT, W. A., WITTSON, C. L.
und HUNT, E. B. A Theoretical and Practical Analysis
of the Diagnostic Process. In Hoch, and
Zubin (Eds.): Current Problems in
Psychiatric Diagnosis.
New York: Grune and Stratton, Inc.
1953

24. KAEHLBLING, R. und
VOLPE, P. A. Constancy of Psychiatric Diagnosis in
Re-admissions.
Comp. Psychiat., 4, 29-38
1963

25. KATZ, M. M., COLE, J. C.
und LOWERY, H. A. Studies of the Diagnostic Process:
The Influence of Symptom Perception,
Past Experience, and Ethnic Back-
ground on Diagnostic Decisions.
Amer. J. Psychiat., 125, 937-947
1969

26. KENDELL, R. E. The Classifications of Depressive
Illness. London, Maudsley Monograph
no. 18
1968

27. KENDELL, R. E. Psychiatric Diagnosis: A Study of how
they are made.
Brit. J. Psychiat., 122, 437-445
1973

28. KENDELL, R. E., COOPER, J. E., Diagnostic Criteria of American and
GOURLAY, A. J., British Psychiatrists.
COPELAND, J. R. M., SHARPE, L. Archives of General Psychiatry, 25,
und GURLAND, B. J. 123-130
1971
29. KENDELL, R. E., EVERETT, B., The Reliability of the "Present State"
COOPER, J. E., Examination.
SARTORIUS, N. und Soc. Psychiat., 3, 123-129
DAVID, M. E. 1968
30. KREITMAN, N., The Reliability of Psychiatric
Diagnosis.
J. Ment. Sci., 107, 876
1961
31. KREITMAN, N., SAINSBURY, P., The Reliability of Psychiatric
MORRISSEY, J., TOWERS, J. Diagnosis: An Analysis.
und SCRIVENER, J. J. Ment. Sci., 107, 887
1961
32. LORR, M., KLETT, C. J., Inpatient Multidimensional Psychiatric
MC. NAIR, D. M. und Scale (IMPS). Palo Alto:
LASKY, J. J. Consulting Psychologists Press
1962
33. MASSERMAN, J. H. und J. Ment. Sci., 84, 893,
CARMICHAEL, H. T. 1938
34. MEHLMAN, B. J. Abnorm. Soc. Psychol. 47, 577
1952
35. MENDEL, B. D. On Therapist-Watching
Psychiatry 27, 59-68
1964
36. NORRIS, V. Mental Illness in London, Maudsley
Monographs, no. 6,
London: Chapman and Hall, Ltd.
1959
37. OVERALL, J. E. und The Brief Psychiatric Rating Scale.
GORHAM, D. R. Psychological Reports, 10, 799-812
1962

37. PASAMANICK, B., DINITZ, S. und LEFTON, M. Amer. J. Psychiat., 116, 127
1959
38. PICHOT, P., BAILLY, R. und OVERALL, J. E. Les stereotypes diagnostiques des psychoses chez les psychiatres Français: comparaisons avec les stereotypes Americains, in Proceedings of the Fourth International Congress of the Collegium Internationale Neuro-psychopharmacologicum. Excerpta Medica International Congress Series, 129, 968-973
1966
39. RAINES, G. N. und ROHRER, J. H. The Operational Matrix of Psychiatric Practice.
Amer. J. Psychiat. 111, 721-733
1954/55
40. RAWNSLEY, K. An International Diagnostic Exercise. In: Proceeding of the Fourth World Congress of Psychiatry, Excerpta Medica Foundation, Amsterdam, 4, 2683-2686
1967
41. SANDIFER, M. G., HORDERN, A. und GREEN, L. M. The Psychiatric Interview: The Impact of the First Three Minutes.
Amer. J. Psychiat., 126, 968-973
1970
42. SANDIFER, M. G., HORDERN, A., TIMBURY, G. C. und GREEN, L. M. Psychiatric Diagnosis: A Comprehensive Study in North Carolina, London and Glasgow.
Brit. J. Psychiat., 114, 1-9
1968
43. SANDIFER, M. G., HORDERN, A., TIMBURY, G. C. und GREEN, L. M. Similarities and Differences in Patient Evaluation by U. S. and U. K. Psychiatrists
Amer. J. Psychiat., 126, 206-212
1969
44. SANDIFER, M. G., PETTUS, CH. und QUADE, D. A Study of Psychiatric Diagnosis.
J. Nerv. Ment. Dis., 139, 350-356
1964

45. SCHMIDT, H. O. und J. Abnorm. Soc. Psychol., 52, 262,
FONDA, C. P. 1956
46. SHARPE, L., GURLAND, B. J., The Accuracy of Transatlantic
FISHER, B. und Communication in Psychiatry.
FLEISS, J. L. Amer. Psychiat. Assoc. Meeting,
Miami, 1969
47. SHEPHERD, M., BROOKE, E. M., An Experimental Approach to
COOPER, J. E. und LIN, T. Psychiatric Diagnosis.
Acta Psychiatrica Scandinavica,
Supplementum 201 ad Volumen 44
1968
48. STENGEL, E. Classification of Mental Disorders,
Bull. Wld. Hlth. Org. 21, 601-663
1959
49. THORNDIKE, E. L. A Constant Error in Psychological
Ratings.
J. Educ. Psychol. 4, 25-29
1920
50. WING, J. K., BIRLEY, J. L. T., Reliability of a Procedure for
COOPER, J. E., GRAHAM, P. Measuring and Classifying "Present
und ISAACS, A. D. Psychiatric State".
Brit. J. Psychiat. 113, 499-515
1967
51. ZUBIN, J. Classification of the Behaviour
Disorders.
Am. Rev. Psychol. 18, 373-401
1967

ANHANG

Die "Einzel diagramme" sämtlicher Beurteiler aus Film 1 und Film 2, aufgeteilt in zwei Gruppen gemäß drei Indices:

1. Flexibilität,
2. Sicherheit,
3. Diagnosen-Summe.

Gruppe 1 entsprechen eine "niedrige Flexibilität", eine "hohe Sicherheit" und eine "niedrige Diagnosen-Summe" (vgl. S. 64).

Gruppe 2 entsprechen eine "hohe Flexibilität", eine "niedrige Sicherheit" und eine "hohe Diagnosen-Summe".

Die Übergänge von Gruppe 1 in Gruppe 2 sind kontinuierlich, die beiden Gruppen lassen sich demnach nicht scharf voneinander abgrenzen.

Zeichenerklärung:

	=	1. diagnostische Hauptgruppe
	=	2. diagnostische Hauptgruppe
	=	3. diagnostische Hauptgruppe
	=	4. diagnostische Hauptgruppe
	=	5. diagnostische Hauptgruppe
	=	6. diagnostische Hauptgruppe

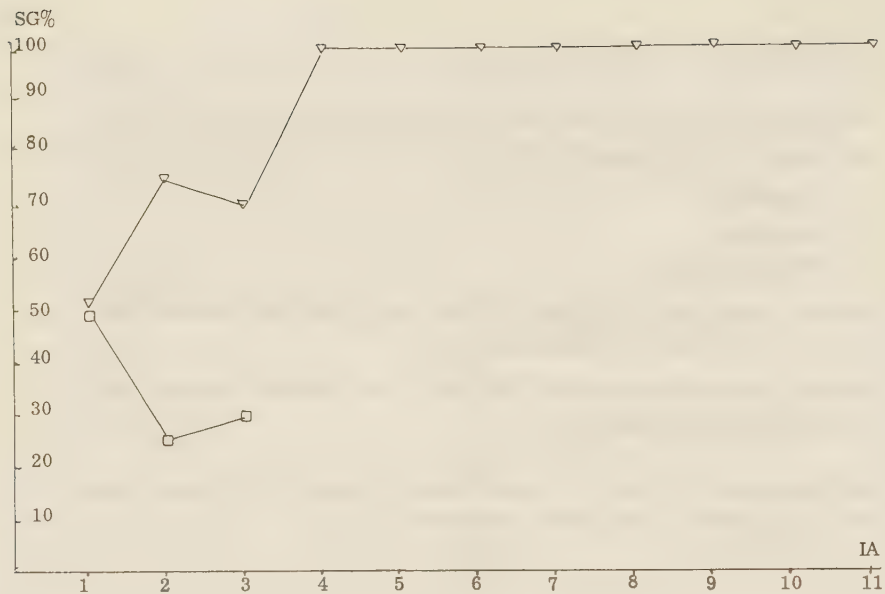


Abb. Nr. 29: Film 1, Gruppe 1.

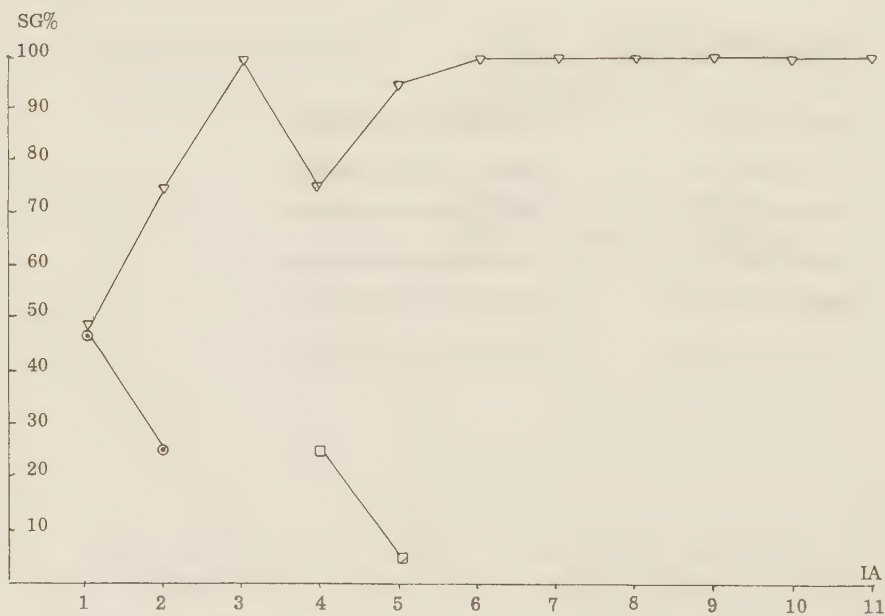


Abb. Nr. 30: Film 1, Gruppe 1.



Abb. Nr. 31: Film 1, Gruppe 1

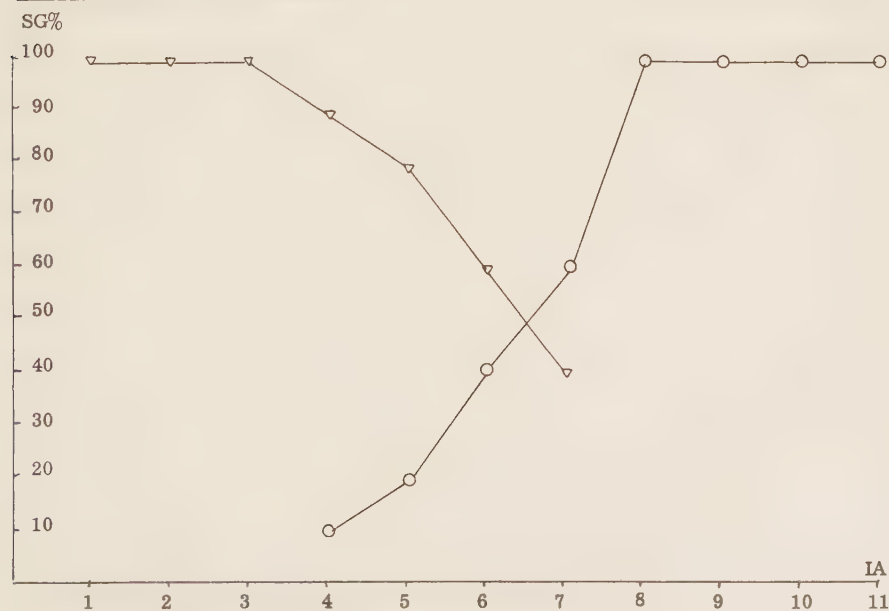


Abb. Nr. 32: Film 1, Gruppe 1.

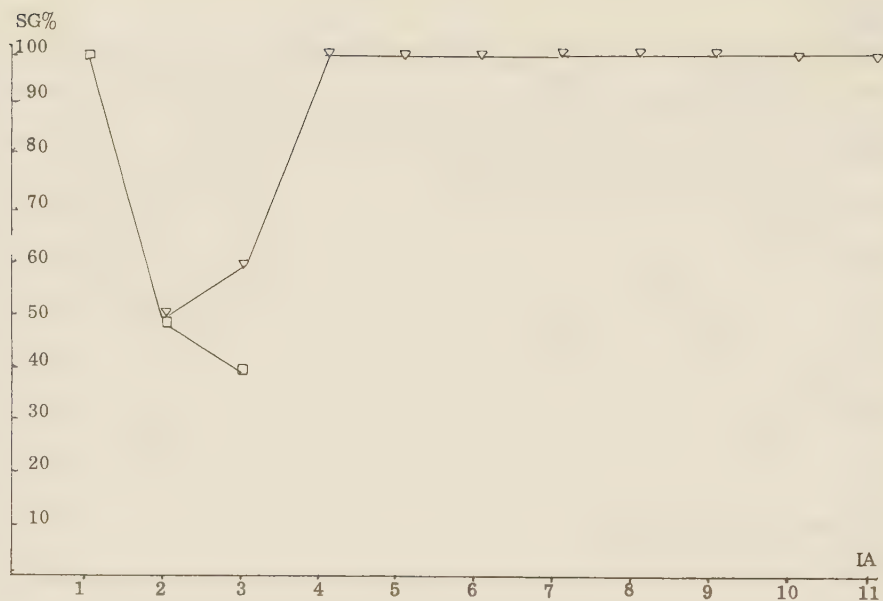


Abb. Nr. 33: Film 1, Gruppe 1.

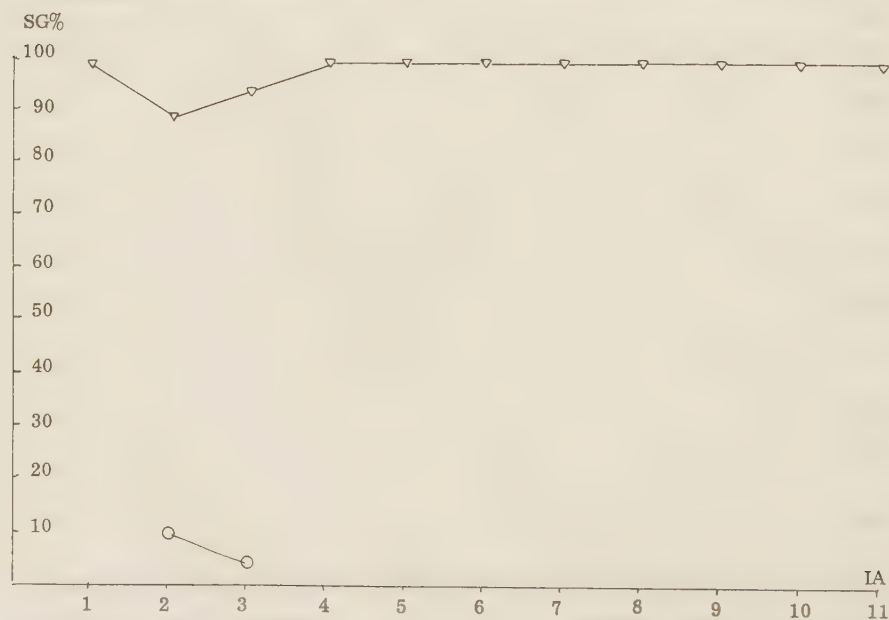


Abb. Nr. 34: Film 1, Gruppe 1.

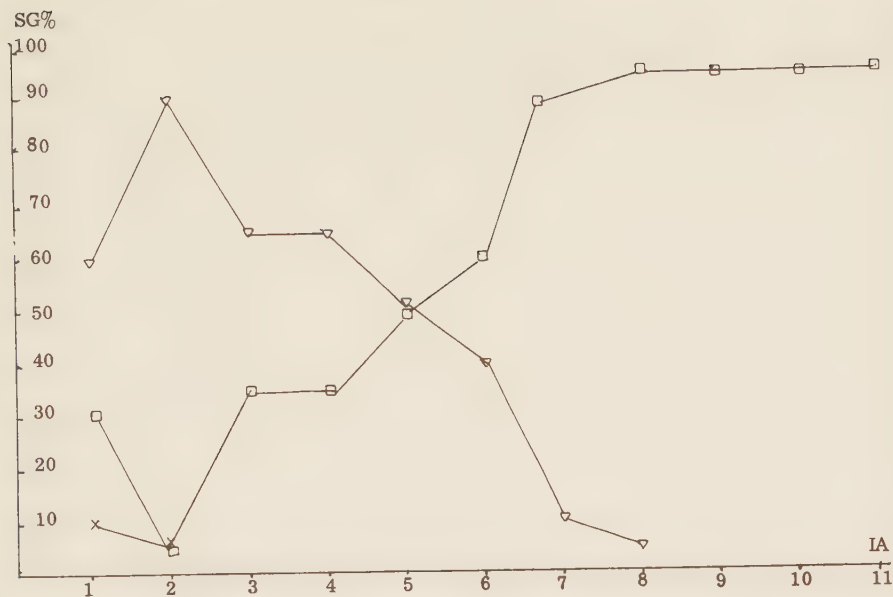


Abb. Nr. 35: Film 1, Gruppe 1.

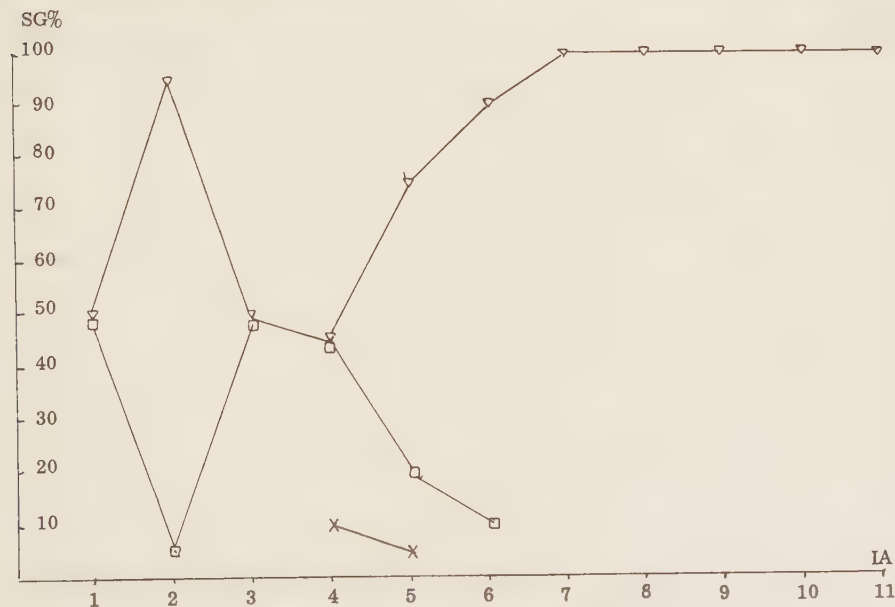


Abb. Nr. 36: Film 1, Gruppe 1.

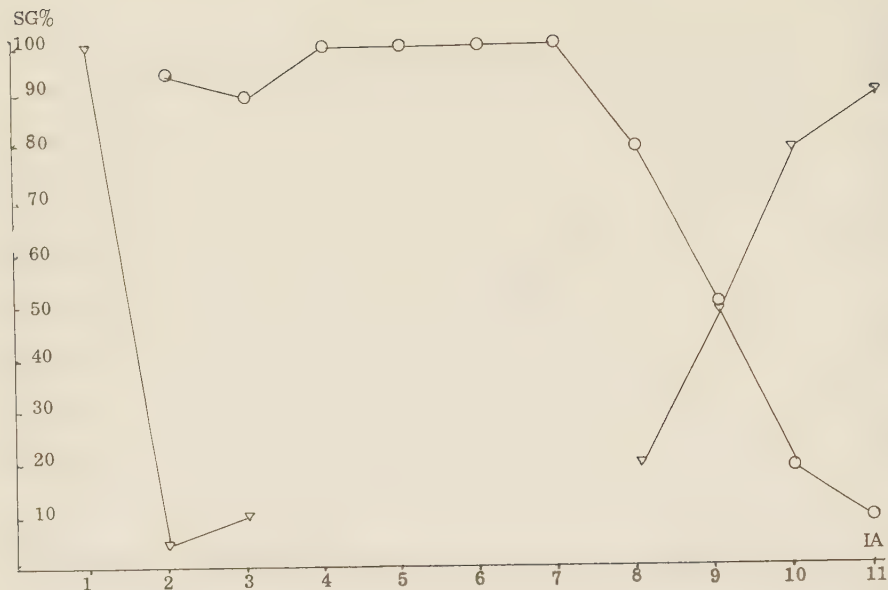


Abb. Nr. 37: Film 1, Gruppe 1.

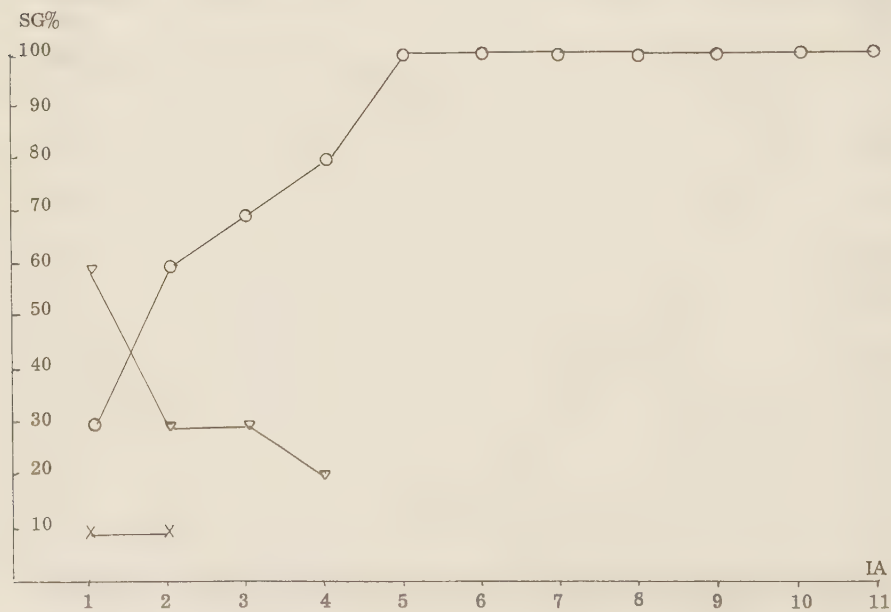


Abb. Nr. 38: Film 1, Gruppe 1.

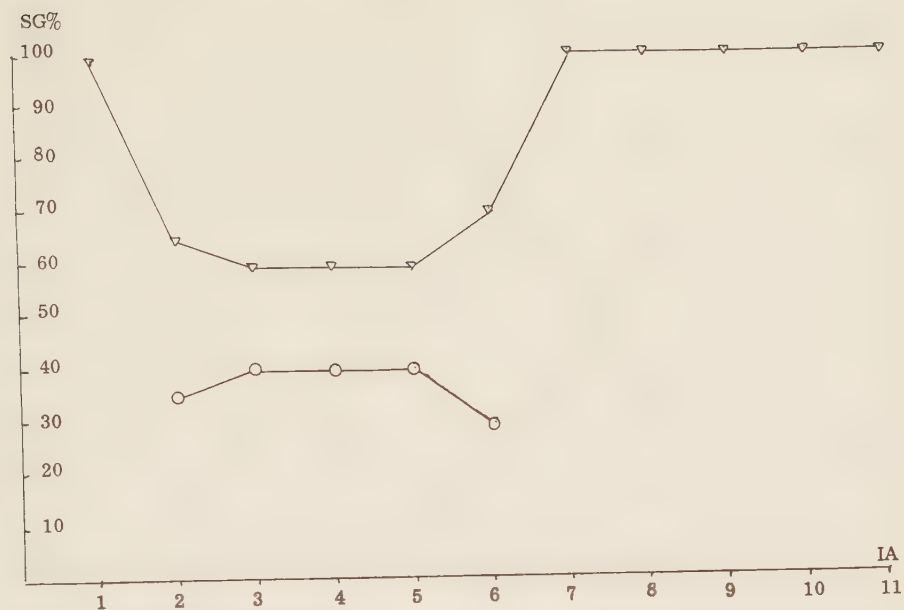


Abb. Nr. 39: Film 1, Gruppe 1.

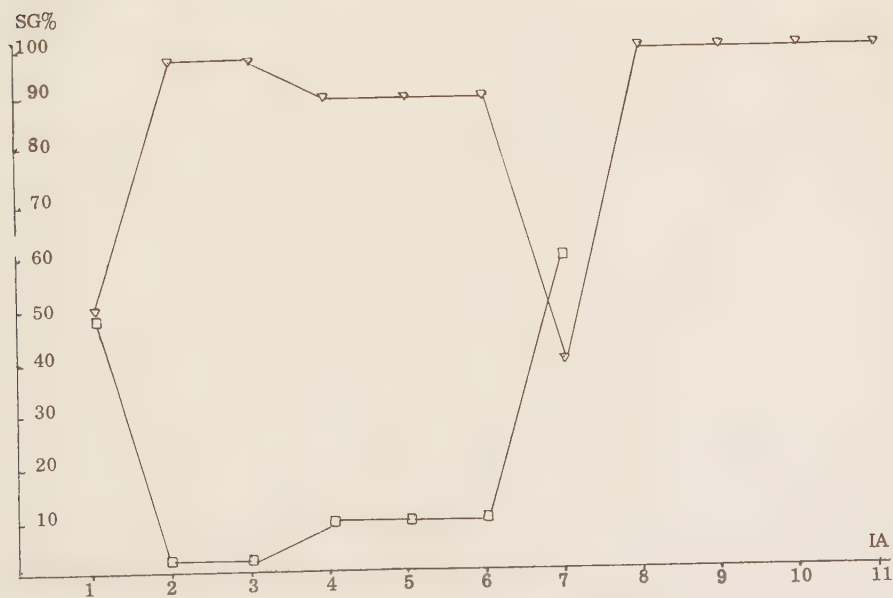


Abb. Nr. 40: Film 1, Gruppe 1.

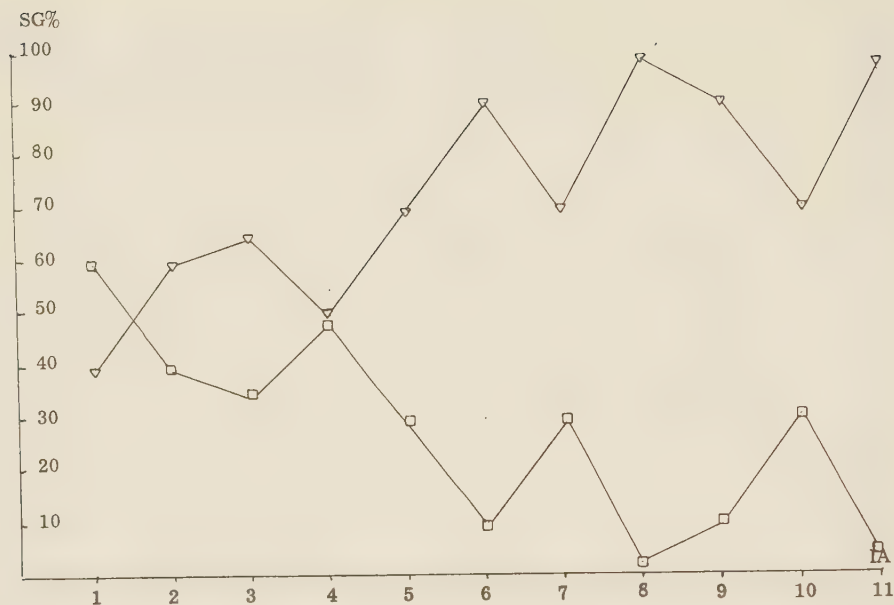


Abb. Nr. 41: Film 1, Gruppe 1.

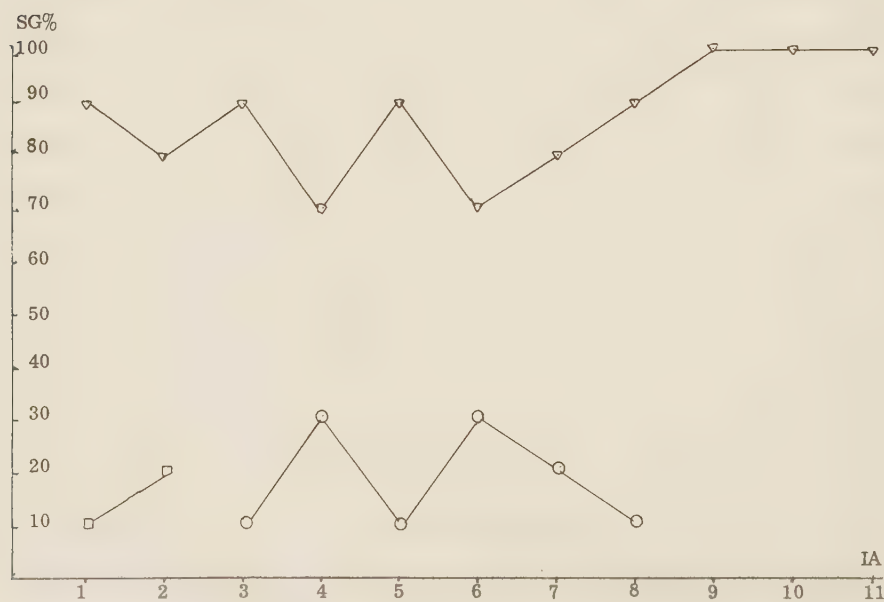


Abb. Nr. 42: Film 1, Gruppe 1.

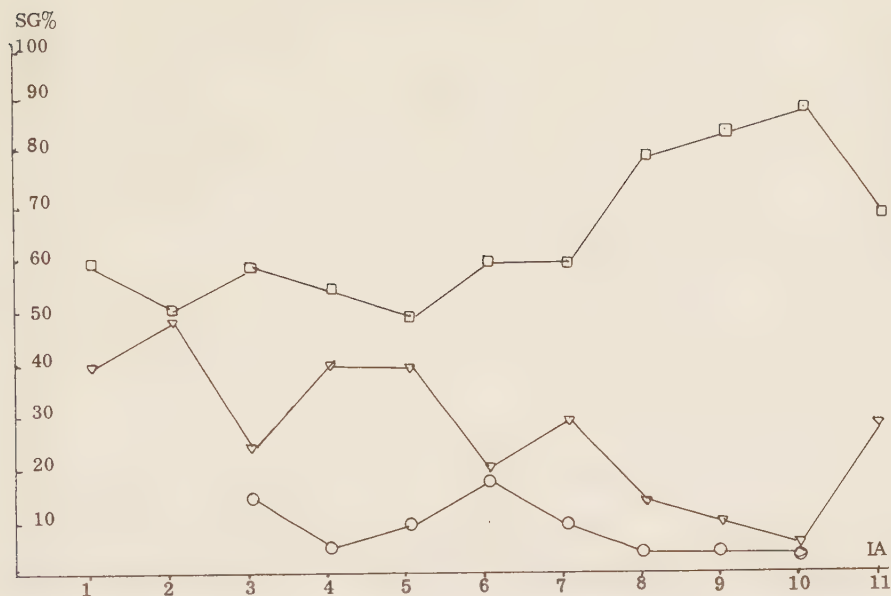


Abb. Nr. 43: Film 1, Gruppe 2.

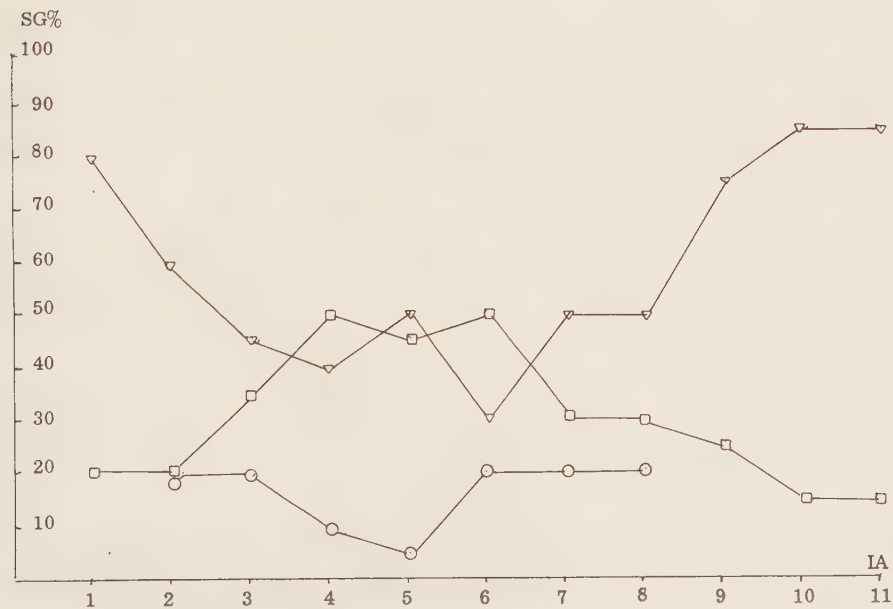


Abb. Nr. 44: Film 1, Gruppe 2.

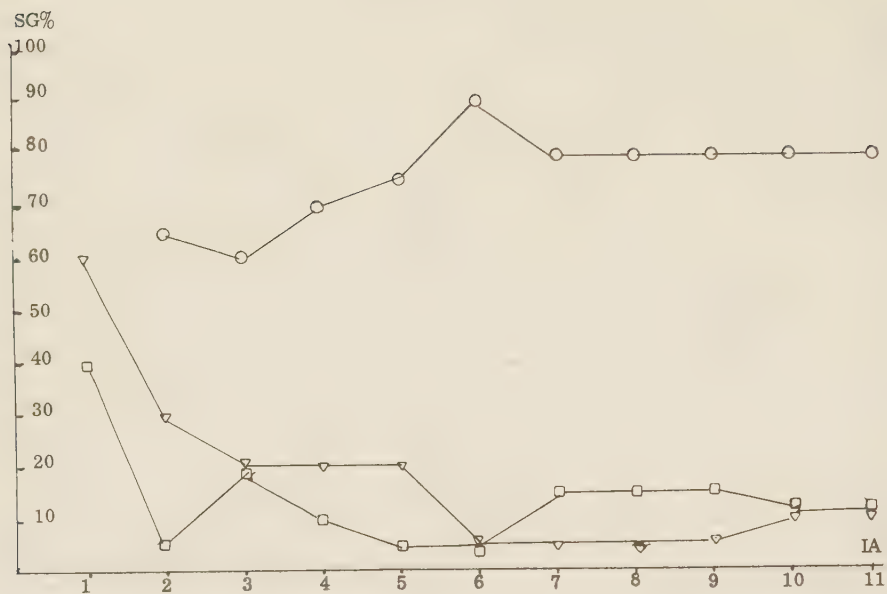


Abb. Nr. 45: Film 1, Gruppe 2.

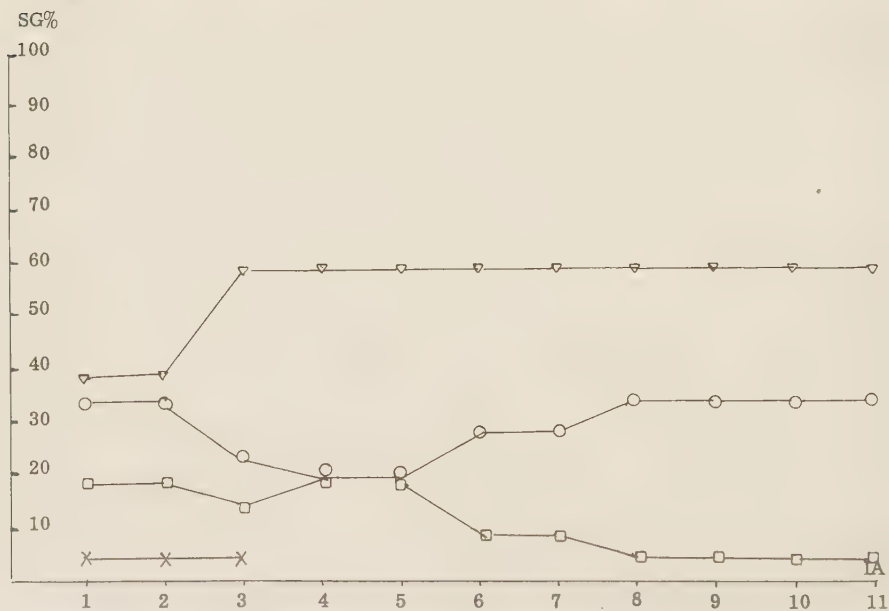


Abb. Nr. 46: Film 1, Gruppe 2.

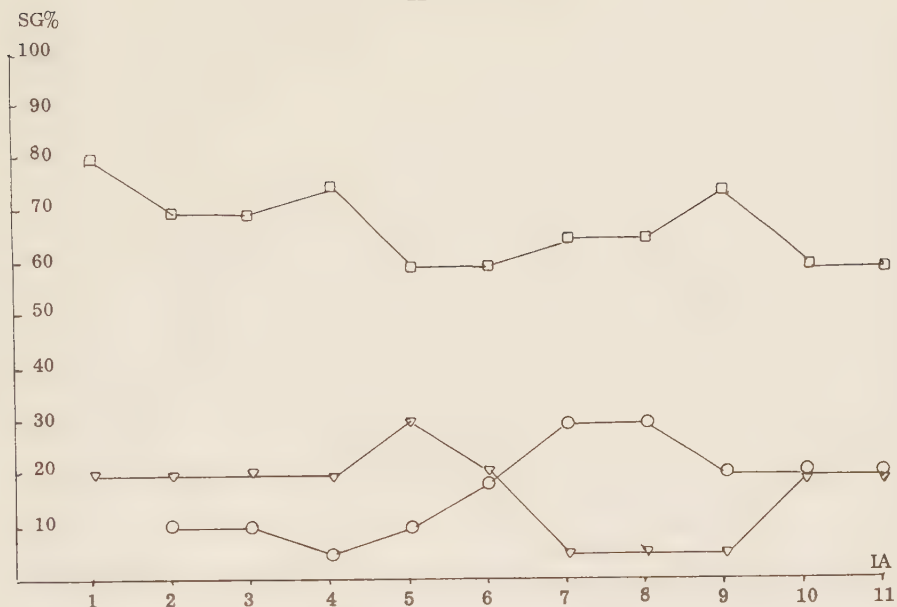


Abb. Nr. 47: Film 1, Gruppe 2.

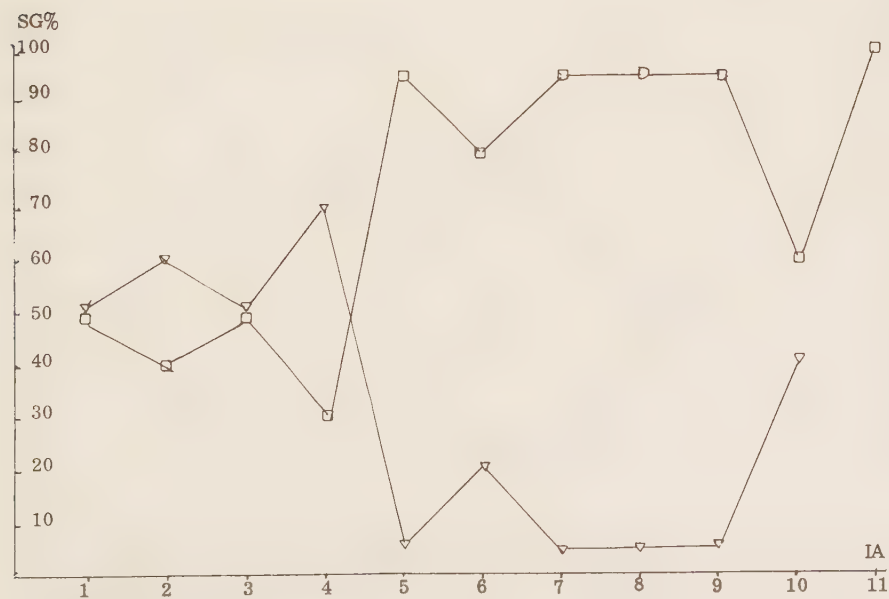


Abb. Nr. 48: Film 1, Gruppe 2.

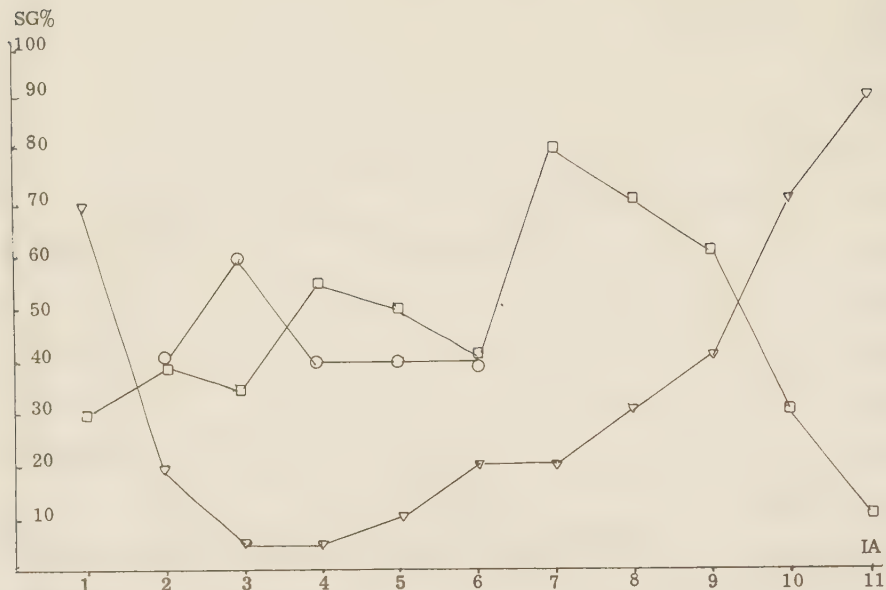


Abb. Nr. 49: Film 1, Gruppe 2.

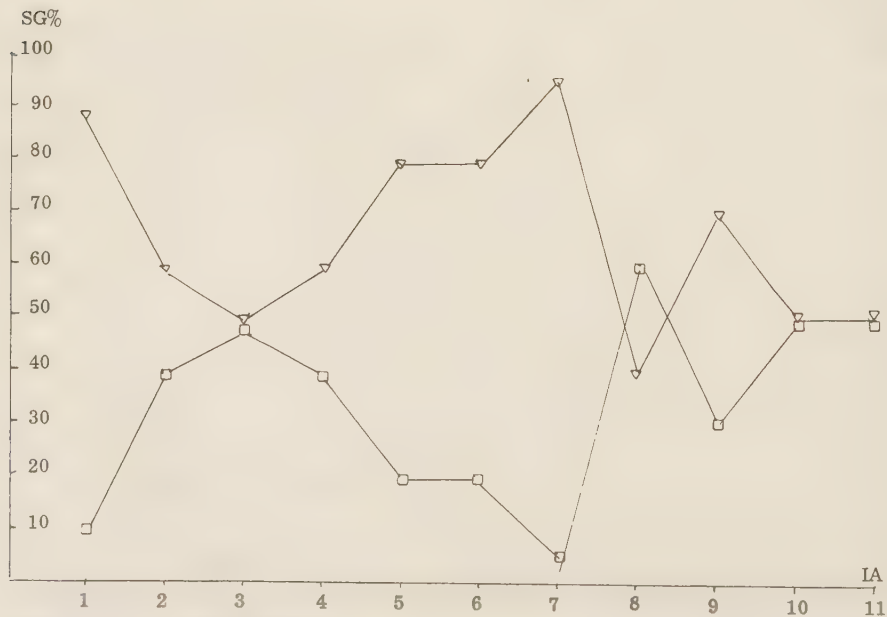


Abb. Nr. 50: Film 1, Gruppe 2.

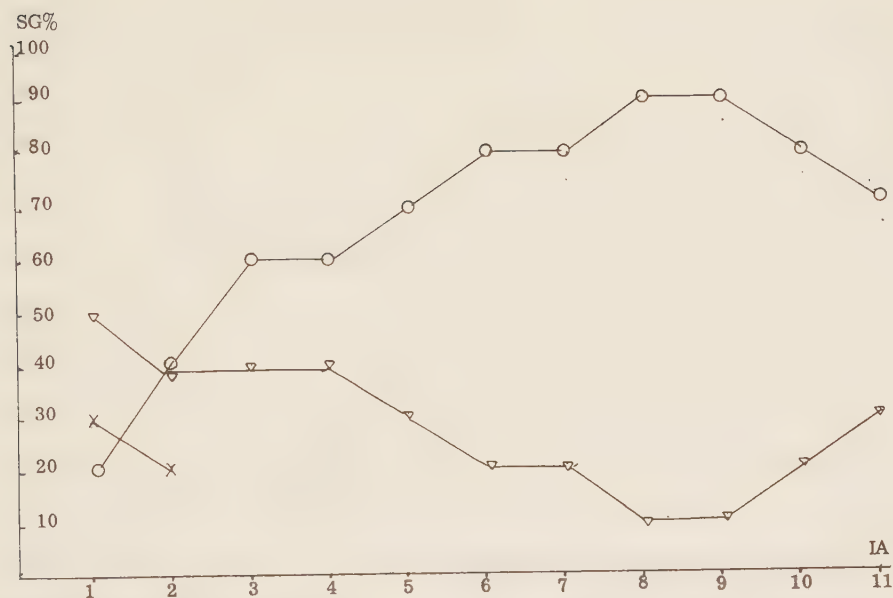


Abb. Nr. 51: Film 1, Gruppe 2.

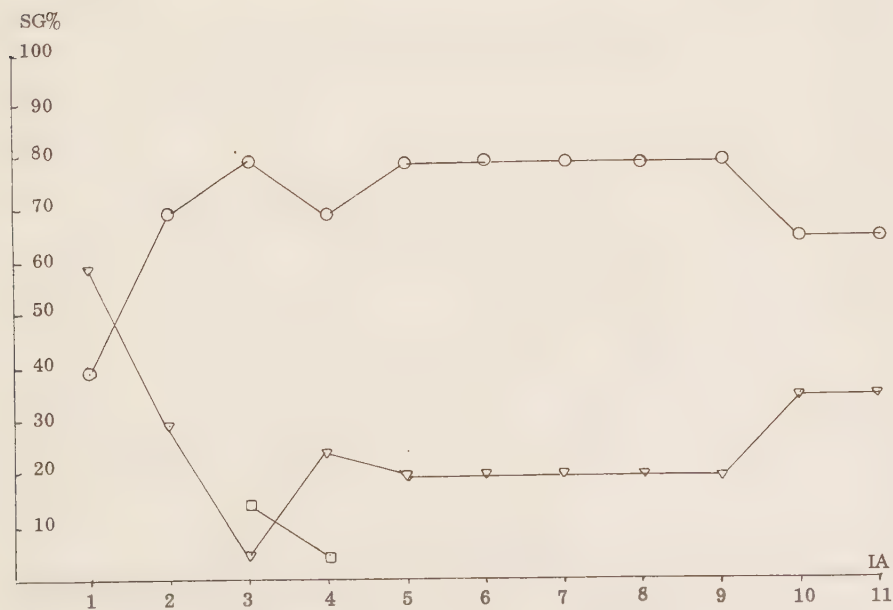


Abb. Nr. 52: Film 1, Gruppe 2.

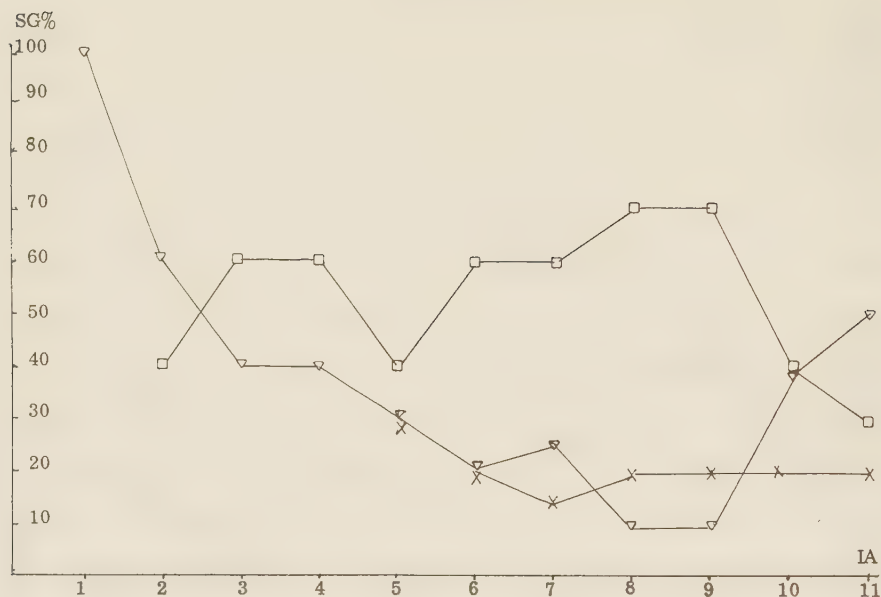


Abb. Nr. 53: Film 1, Gruppe 2.

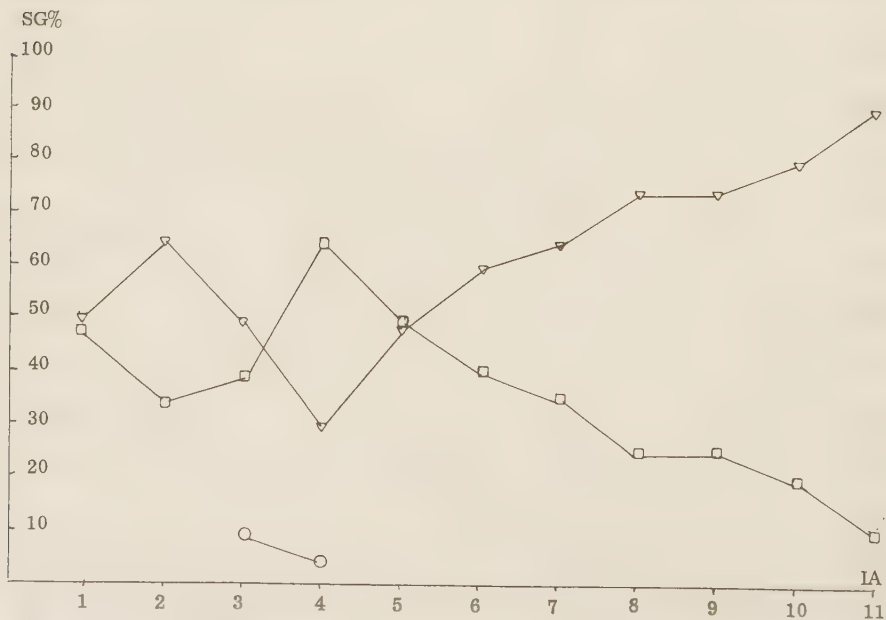


Abb. Nr. 54: Film 1, Gruppe 2.



Abb. Nr. 55: Film 1, Gruppe 2.

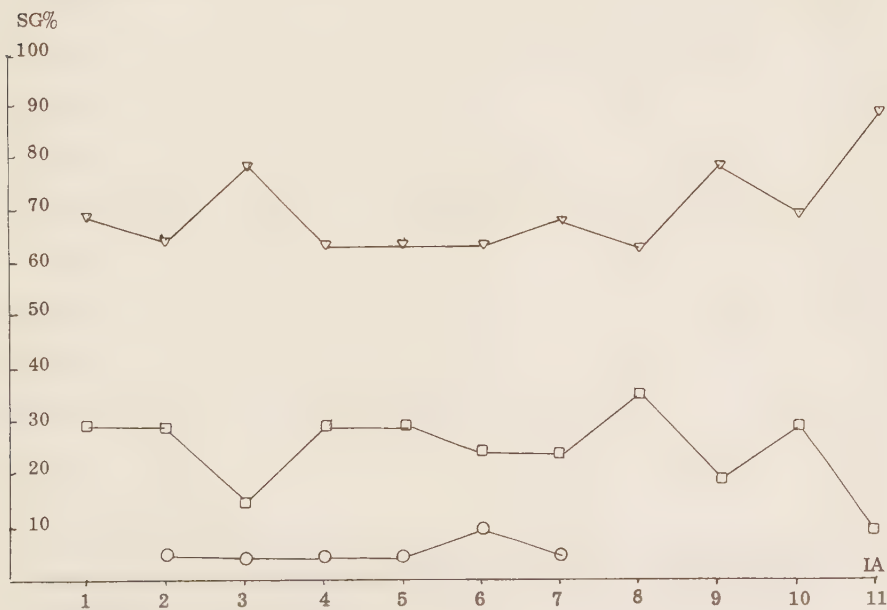


Abb. Nr. 56: Film 1, Gruppe 2.

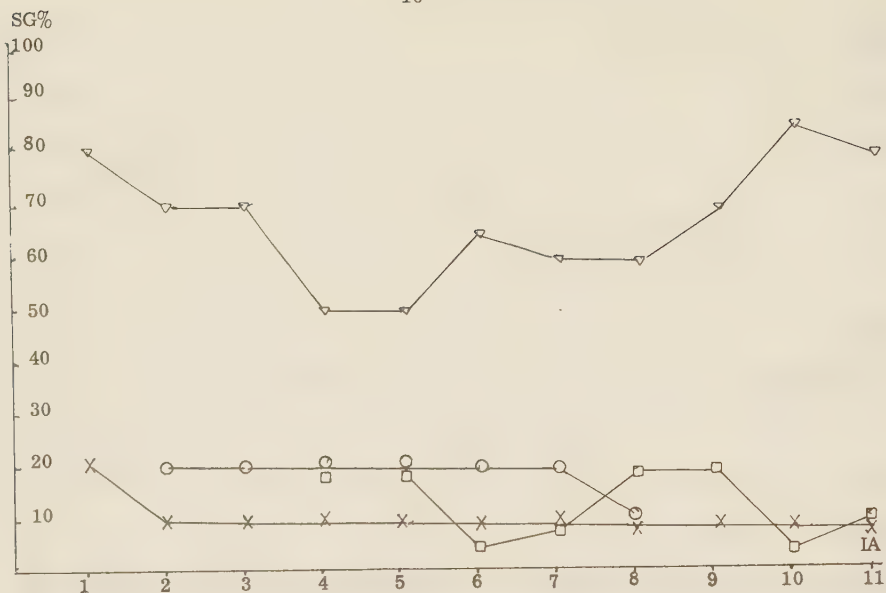


Abb. Nr. 57: Film 1, Gruppe 2.



Abb. Nr. 58: Film 1, Gruppe 2.

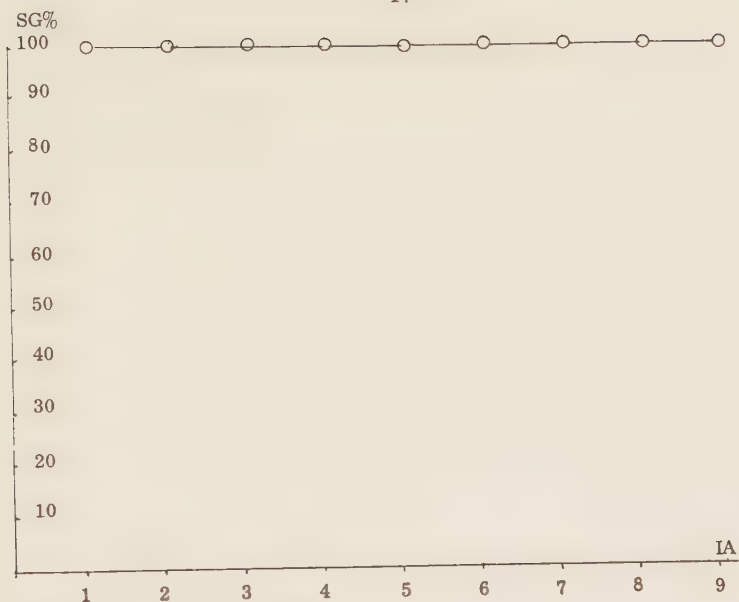


Abb. Nr. 59: Film 2, Gruppe 1.

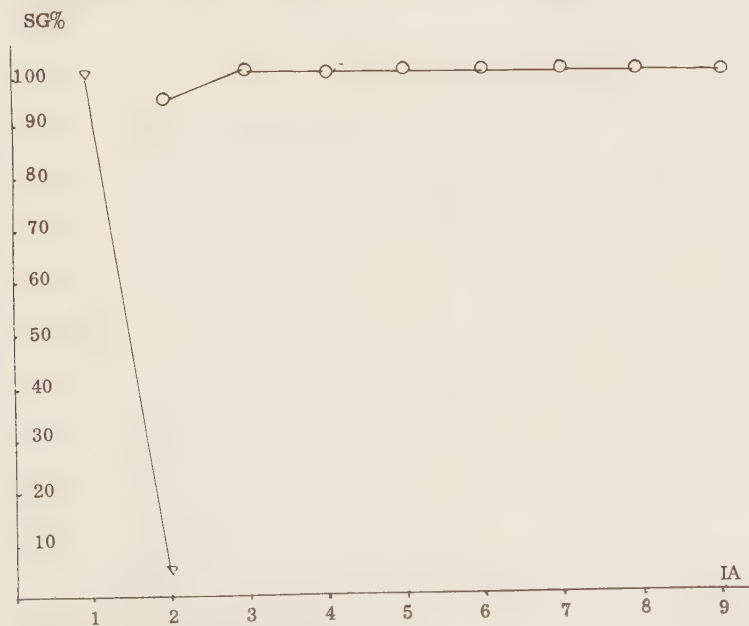


Abb. Nr. 60: Film 2, Gruppe 1.

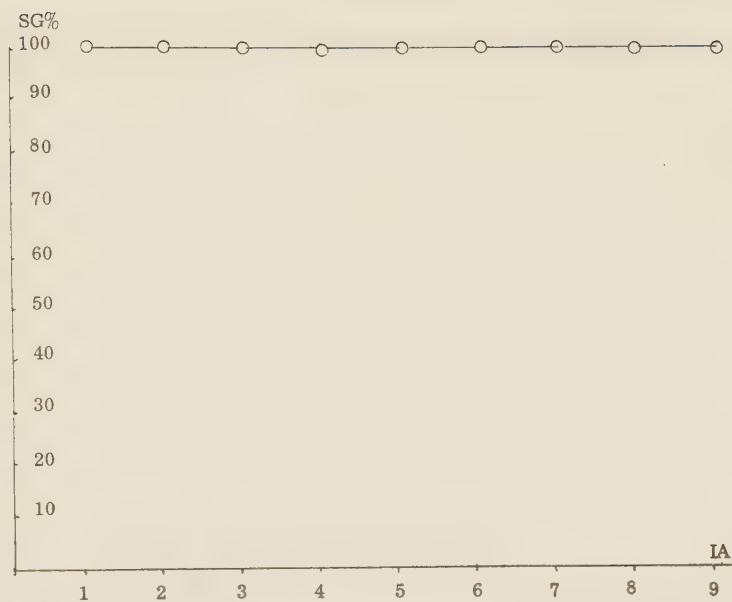


Abb. Nr. 61: Film 2, Gruppe 1.



Abb. Nr. 62: Film 2, Gruppe 1

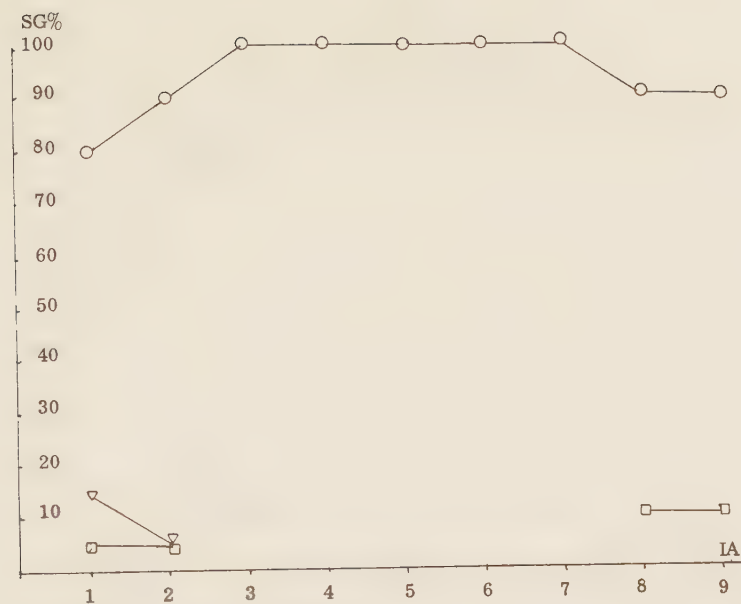


Abb. Nr. 63: Film 2, Gruppe 1.

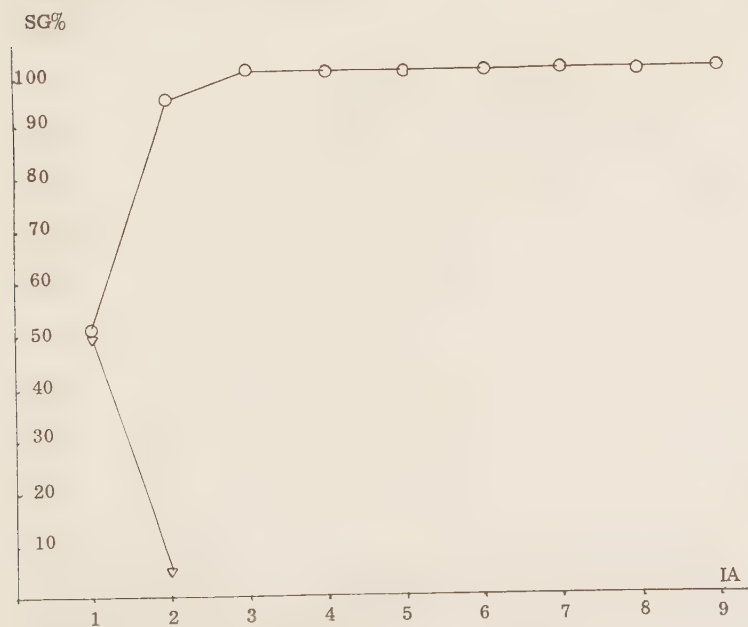


Abb. Nr. 64: Film 2, Gruppe 1.

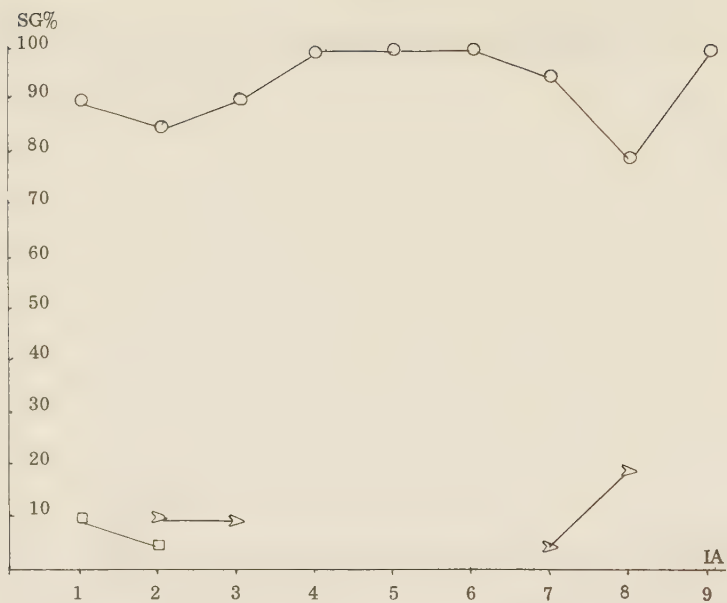


Abb. Nr. 65: Film 2, Gruppe 1.

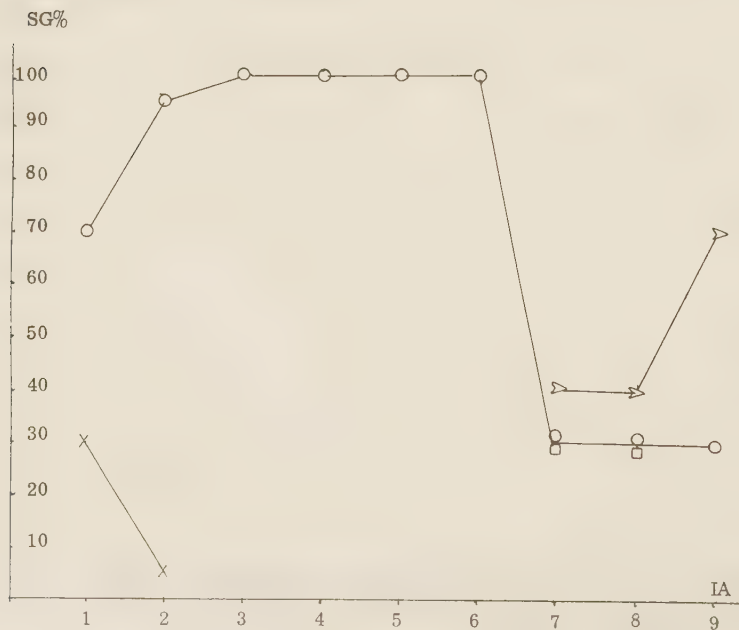


Abb. Nr. 66: Film 2, Gruppe 1.

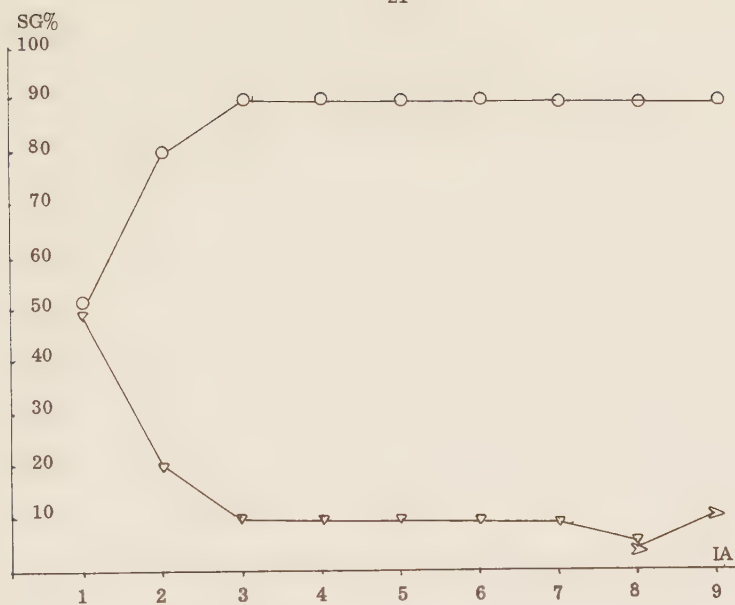


Abb. Nr. 67: Film 2, Gruppe 1.

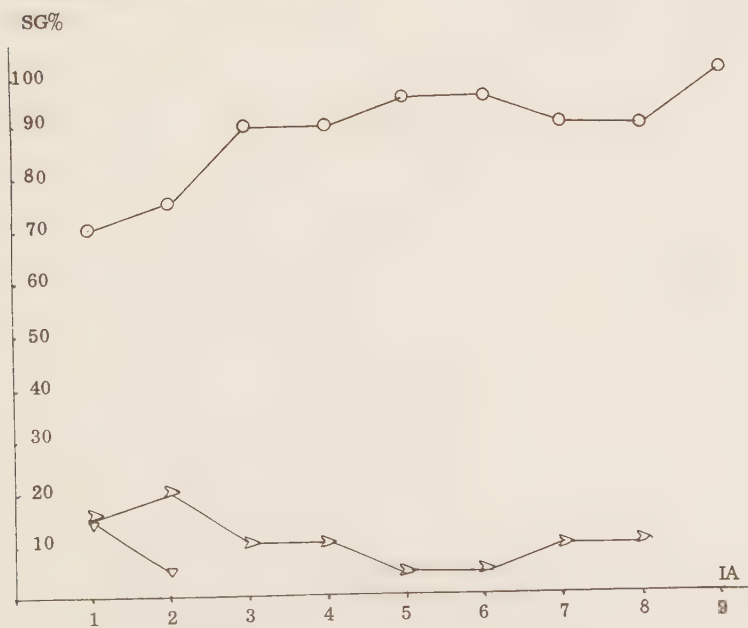


Abb. Nr. 68: Film 2, Gruppe 1.

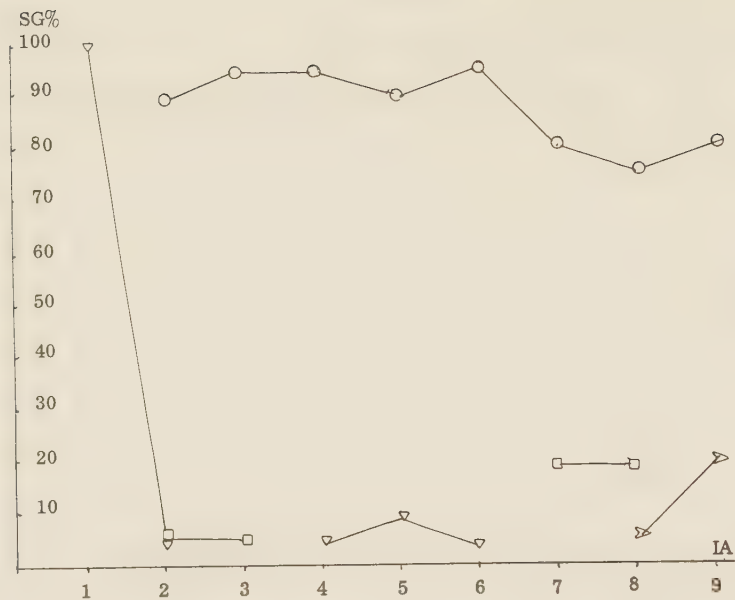


Abb. Nr. 69: Film 2, Gruppe 1.

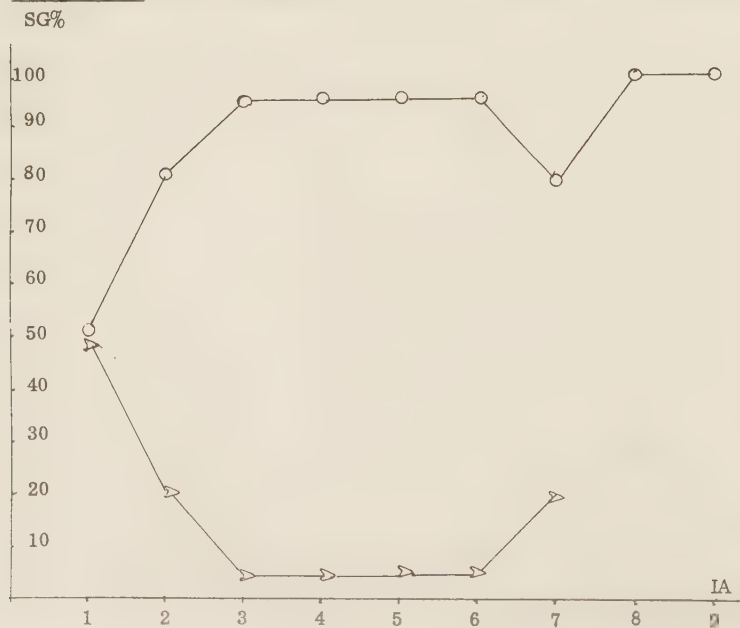


Abb. Nr. 70: Film 2, Gruppe 1.

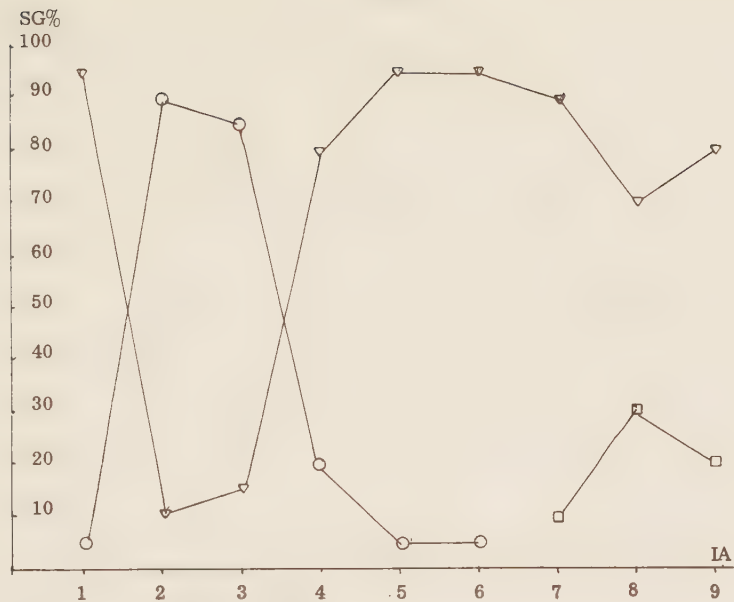


Abb. Nr. 71: Film 2, Gruppe 1.

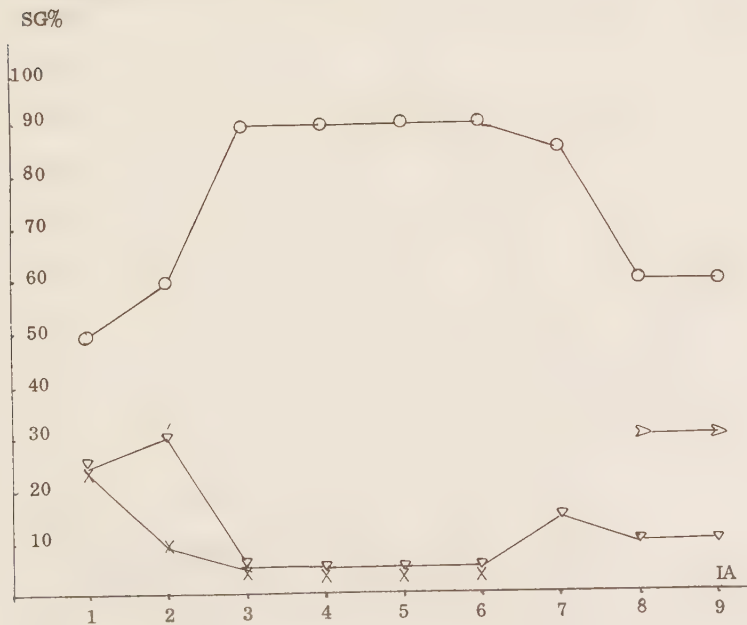


Abb. Nr. 72: Film 2, Gruppe 1.

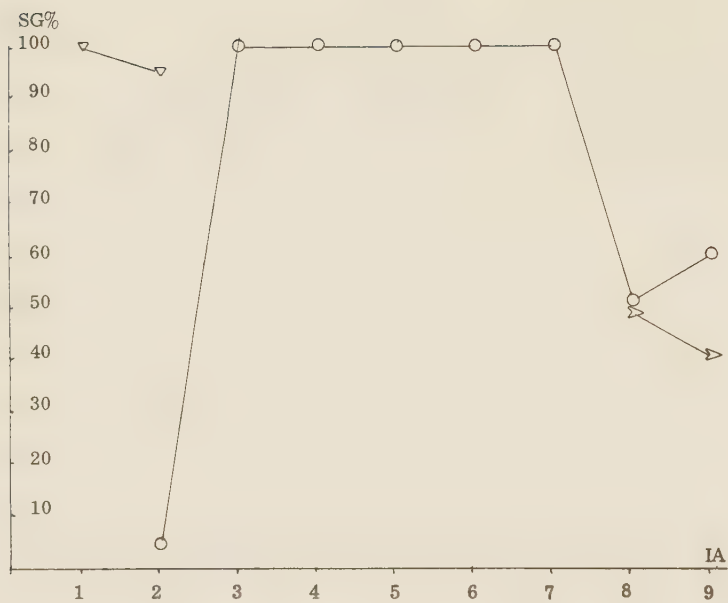


Abb. Nr. 73: Film 2, Gruppe 1.

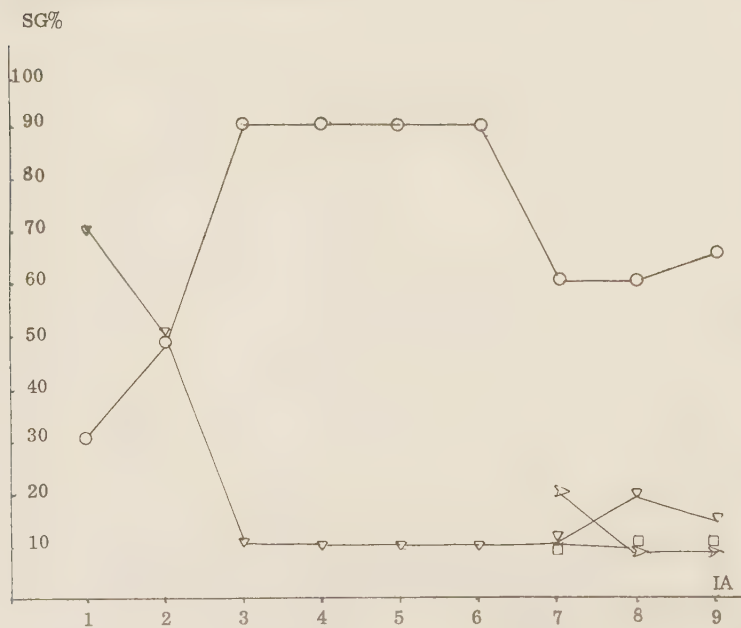


Abb. Nr. 74: Film 2, Gruppe 1.

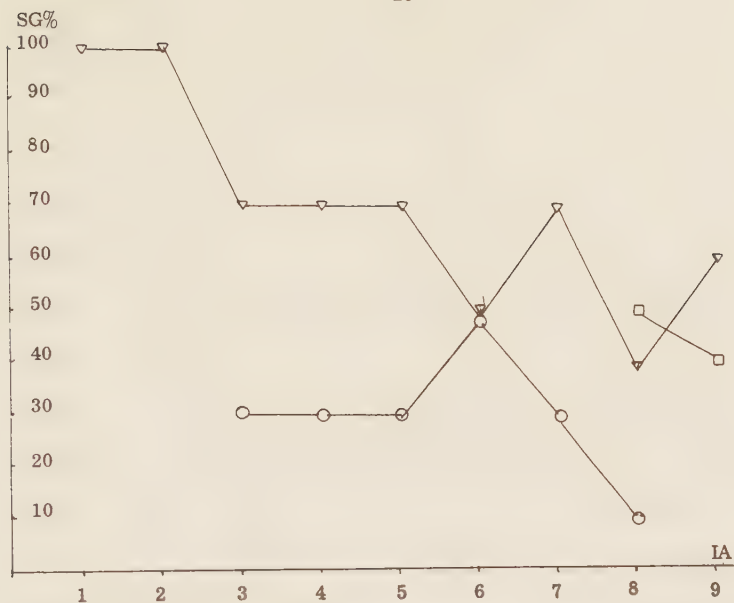


Abb. Nr. 75: Film 2, Gruppe 1.

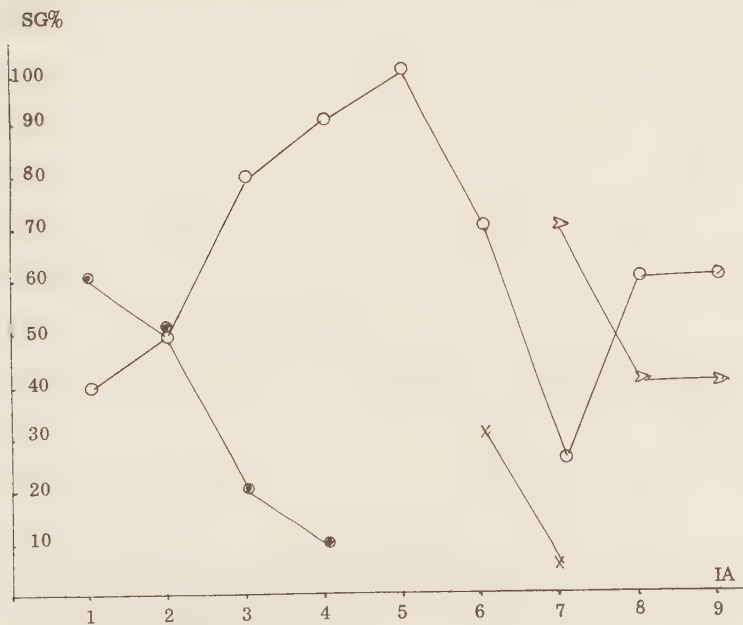


Abb. Nr. 76: Film 2, Gruppe 1.



Abb. Nr. 77: Film 2, Gruppe 2.

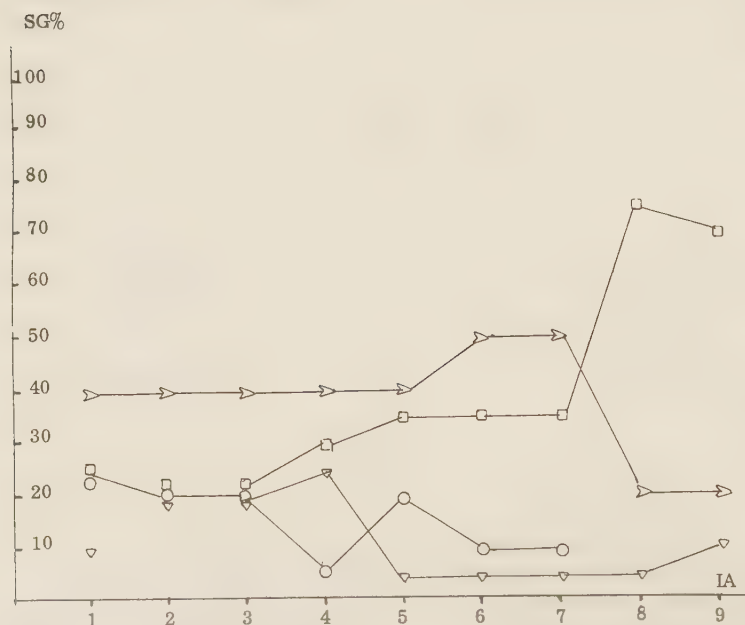


Abb. Nr. 78: Film 2, Gruppe 2.

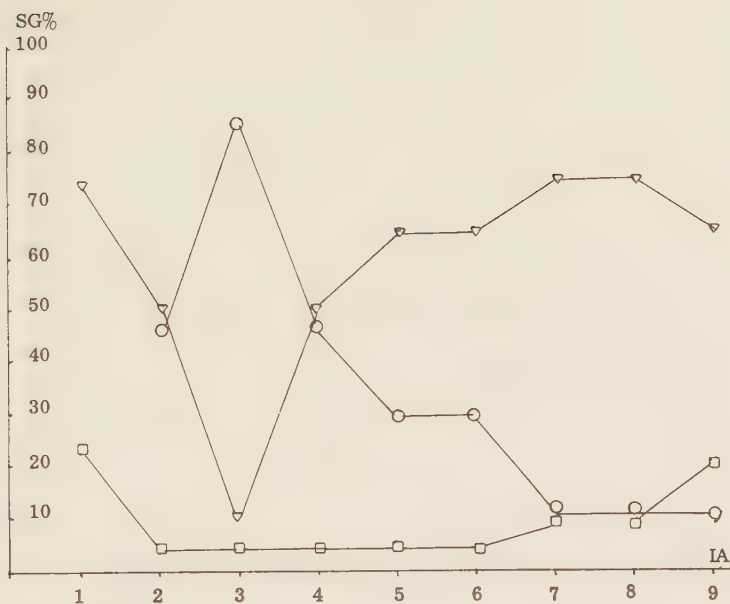


Abb. Nr. 79: Film 2, Gruppe 2.

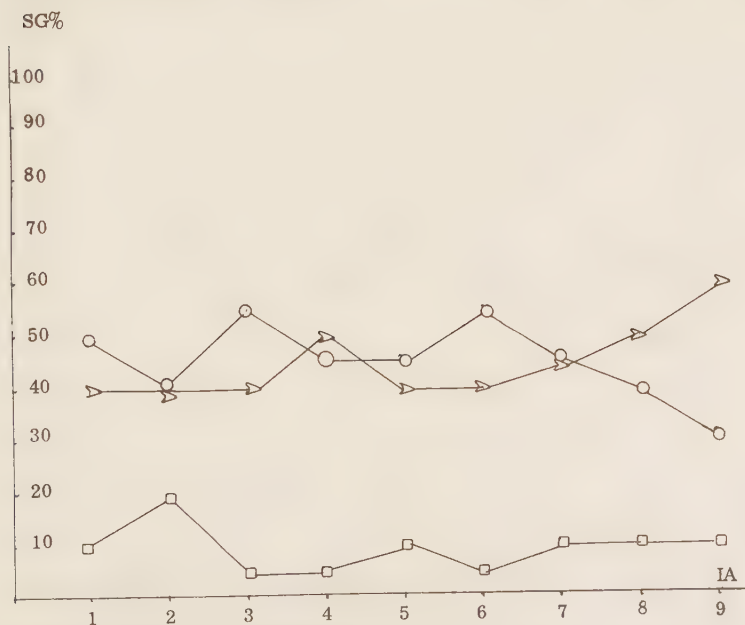


Abb. Nr. 80: Film 2, Gruppe 2.

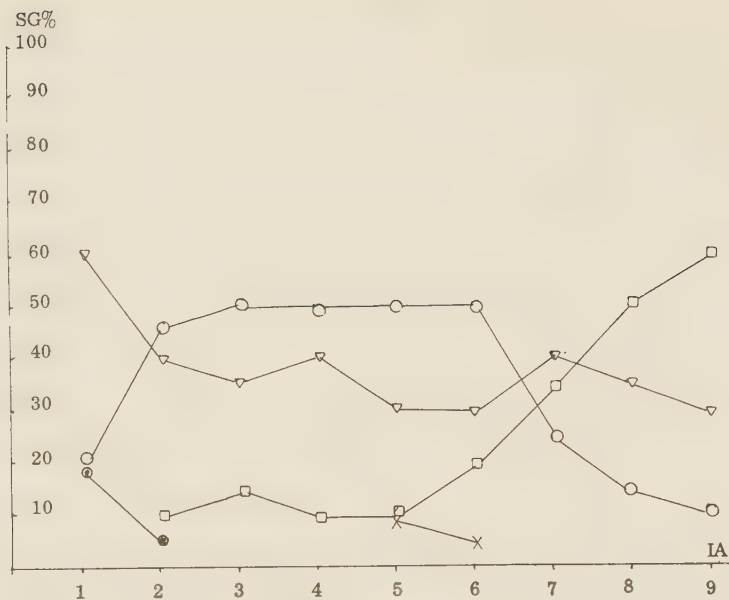


Abb. Nr. 81: Film 2, Gruppe 2.

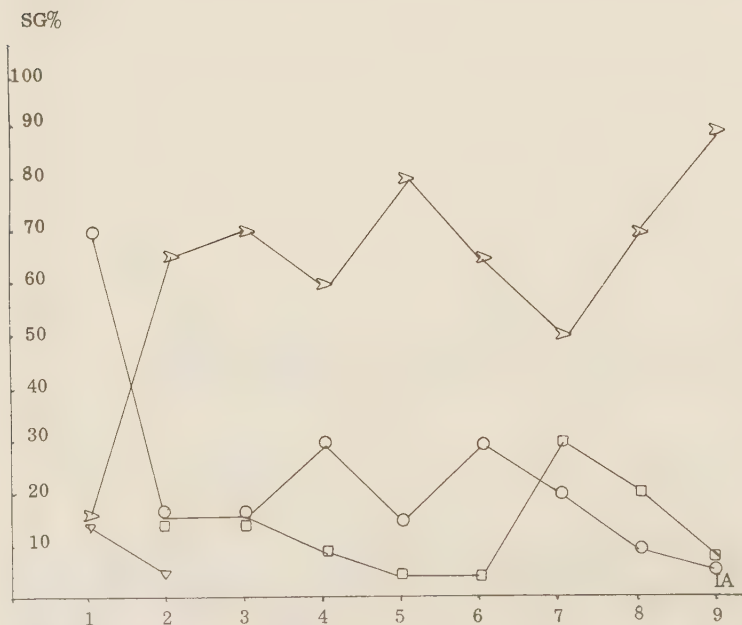


Abb. Nr. 82: Film 2, Gruppe 2.

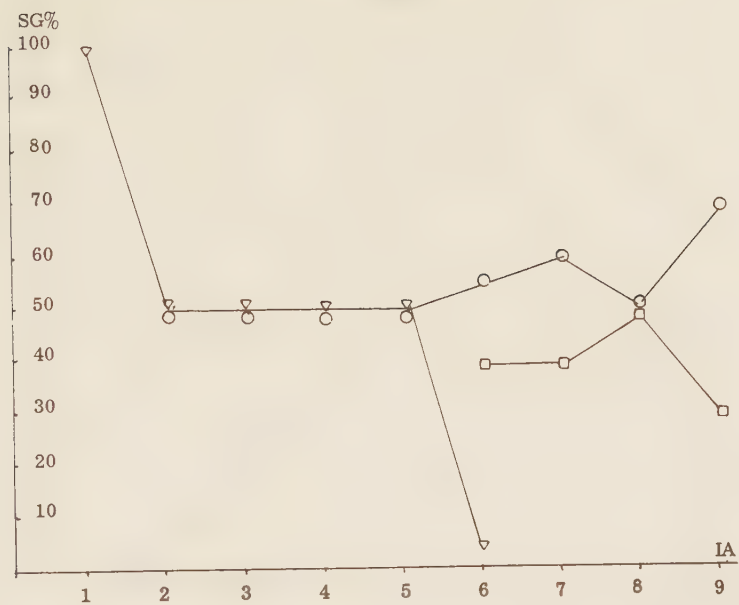


Abb. Nr. 83: Film 2, Gruppe 2.

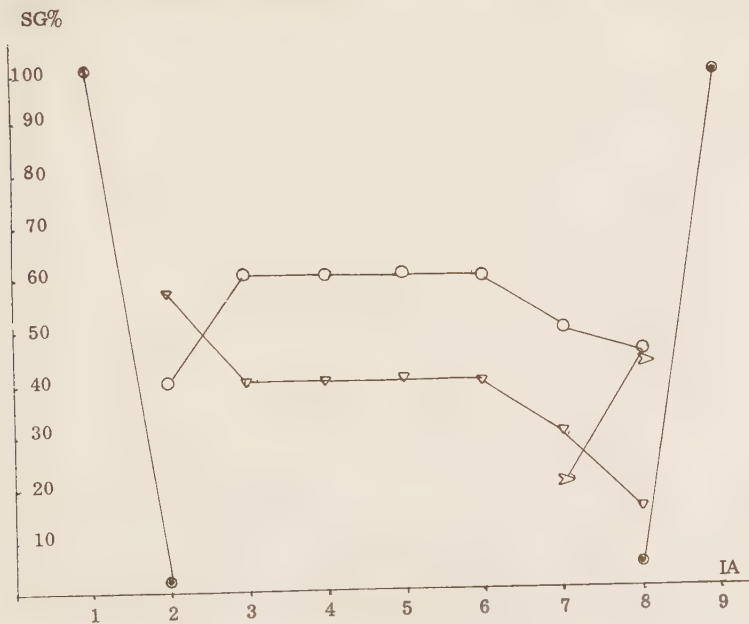


Abb. Nr. 84: Film 2, Gruppe 2.

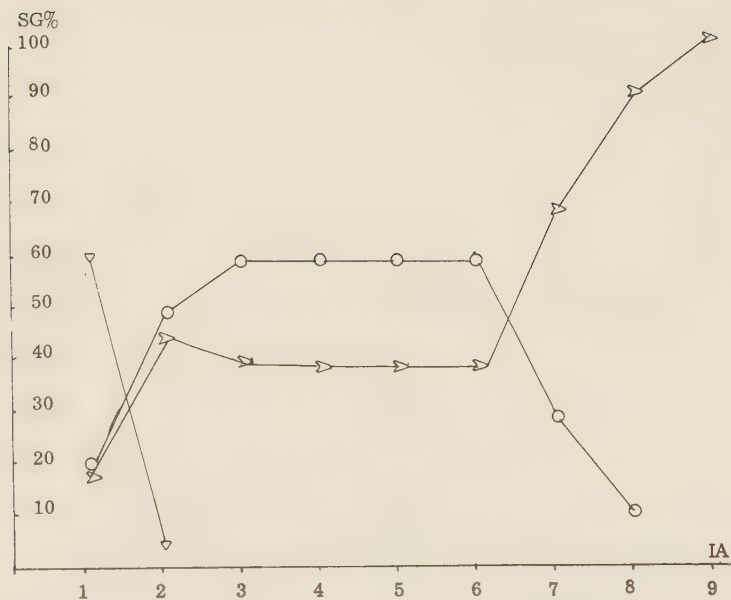


Abb. Nr. 85: Film 2, Gruppe 2.

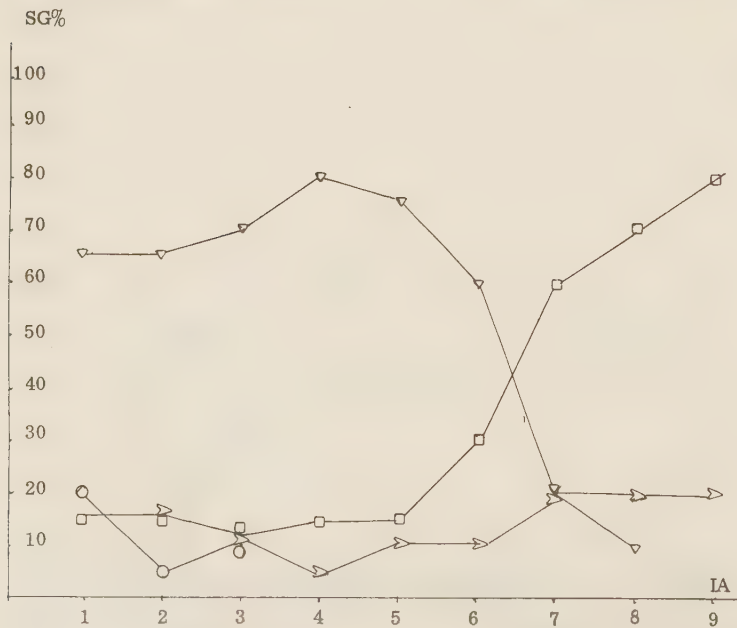


Abb. Nr. 86: Film 2, Gruppe 2.

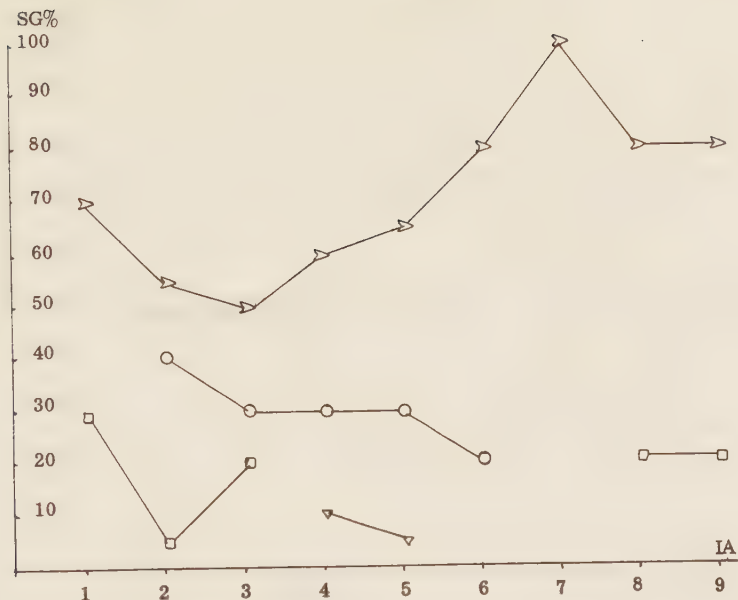


Abb. Nr. 87: Film 2, Gruppe 2.

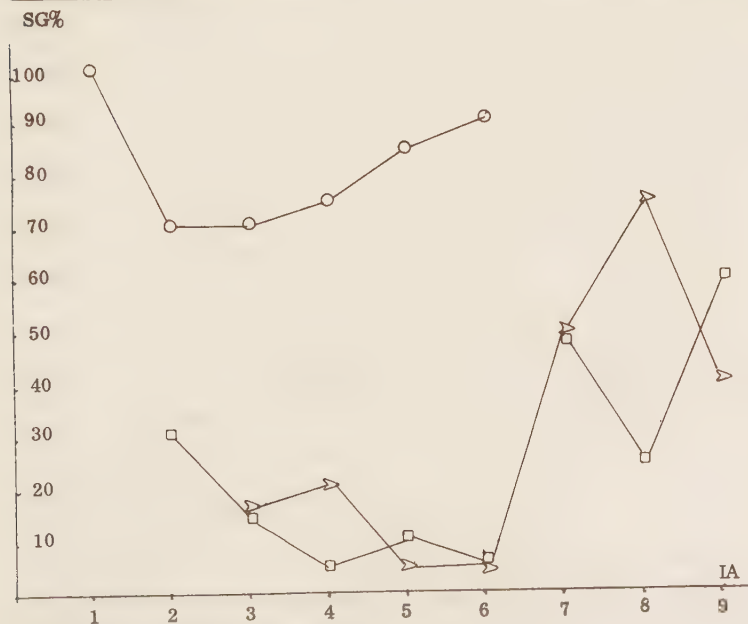


Abb. Nr. 88: Film 2, Gruppe 2.

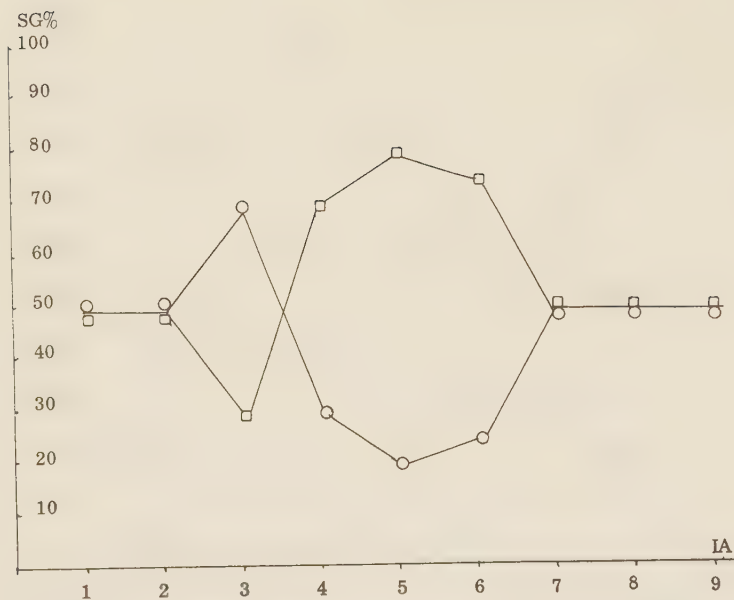


Abb. Nr. 89: Film 2, Gruppe 2.

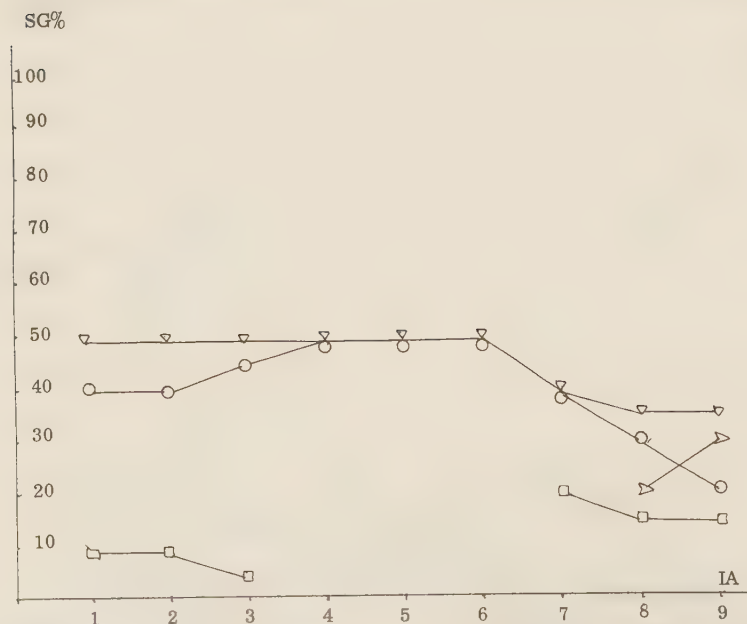


Abb. Nr. 90: Film 2, Gruppe 2.

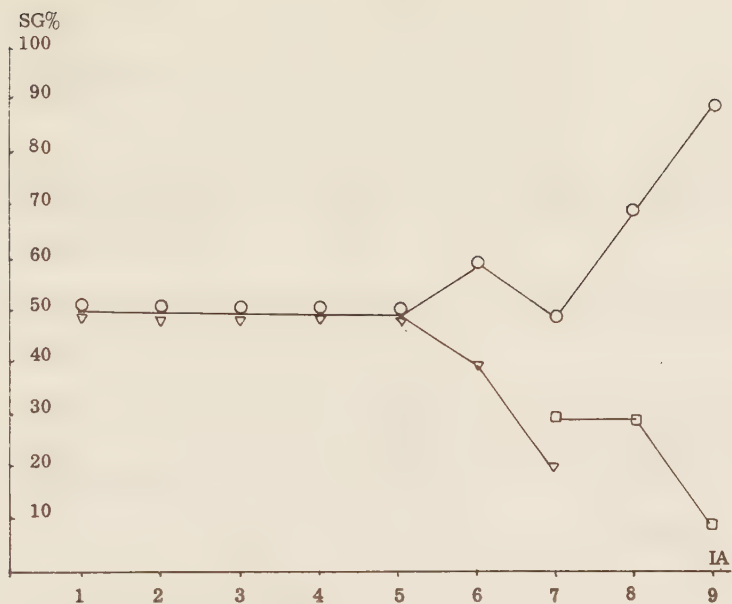


Abb. Nr. 91: Film 2, Gruppe 2.

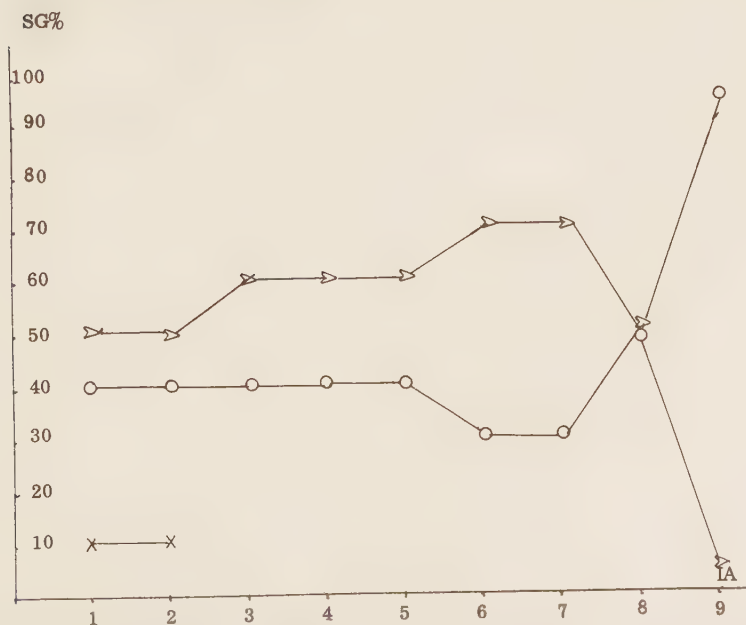
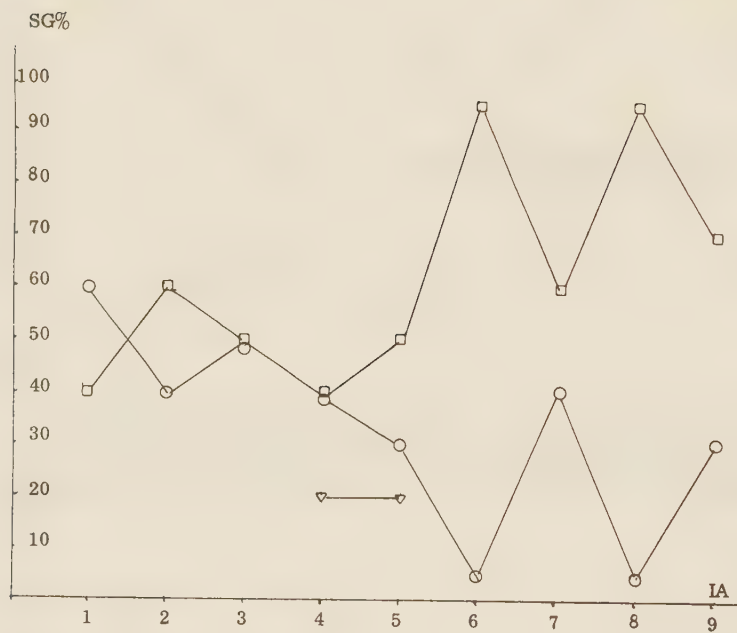


Abb. Nr. 92: Film 2, Gruppe 2.



LEBENS LAUF

Am 18. Mai 1946 wurde ich, Christine Tabbert-Haugg, geb. Haugg, als Tochter des Betriebsleiters Albert Haugg und seiner Ehefrau Irmgard, geb. Weber, in München geboren.

Ab 1957 besuchte ich das Theodolinden-Gymnasium in München-Harlaching, wo ich im Juli 1966 die Reifeprüfung ablegte.

Mein Medizinstudium begann ich im Herbst 1966 an der Ludwigs-Maximilians-Universität. Hier bestand ich im Oktober 1969 die Ärztliche Vorprüfung, im Dezember 1972 schloß ich mein Medizinstudium mit dem Staatsexamen ab.

Im August 1973 heiratete ich den Assistenzarzt Dr. med. Klaus Tabbert..

Meine Medizinalassistentenzeit leistete ich in den Fächern Psychiatrie, Innere Medizin und Chirurgie ab. Am 31. 12. 1974 erhielt ich die Approbation als Ärztin.

